GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: December 14, 2002, 15:45:49 ; Search time 30.5 Seconds
(without alignments)
25.216 Million cell updates/sec

Title: US-09-726-470A-35
Perfect score: 41
Sequence: 1 HAKRRLIF 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	92.7	47	2	S39358	cyclin kinase inhi
2	38	92.7	181	2	I54380	cyclin-dependent k
3	38	92.7	181	2	I68674	cyclin-dependent k
4	37	90.2	159	2	I49023	tumor suppressor p
5	37	90.2	164	2	I84725	tumor suppressor p
6	35	85.4	531	2	S41986	26S proteasome reg
7	33	80.5	214	2	H89959	conserved hypothet
8	31	75.6	76	2	E96619	protein T30E16.11
9	31	75.6	219	2	AG2036	hypothetical prote
10	31	75.6	294	2	S60545	envelope polyprote
11	31	75.6	294	2	S60524	envelope polyprote
12	31	75.6	299	1	A38705	heterocyst develop
13	31	75.6	299	2	S60529	envelope polyprote
14	31	75.6	299	2	S60552	envelope polyprote
15	31	75.6	299	2	S60553	envelope polyprote
16	31	75.6	299	2	S60554	envelope polyprote
17	31	75.6	299	2	AD2098	heterocyst differe
18	31	75.6	300	2	S60547	envelope polyprote
19	31	75.6	300	2	S60546	envelope polyprote
20	31	75.6	300	2	S60556	envelope polyprote
21	31	75.6	300	2	S60555	envelope polyprote
22	31	75.6	300	2	S60557	envelope polyprote
23	31	75.6	301	2	S60548	envelope polyprote
24	31	75.6	302	2	AI0578	citrate (pro-3S)-l
25	31	75.6	303	2	S60550	envelope polyprote
26	31	75.6	303	2	S60549	envelope polyprote
27	31	75.6	311	2	H95877	hypothetical prote
28	31	75.6	377	2	B97757	hypothetical prote
29	31	75.6	487	1	VZEBPT	sensor kinase phoQ
30	31	75.6	487	2	AG0646	sensor protein Pho
31	31	75.6	688	2	E86409	hypothetical prote
32	30	73.2	102	2	S70309	hypothetical prote
33	30	73.2	141	2	S30832	hypothetical prote
34	30	73.2	210	2	S76957	hypothetical prote
35	30	73.2	242	2	F97158	hypothetical prote
36	30	73.2	250	2	B71321	conserved hypothet
37	30	73.2	458	2	F86433	protein T17H7.5 [i
38	30	73.2	751	2	S38101	hypothetical prote
39	30	73.2	891	2	T38195	probable alpha,alp
40	30	73.2	4848	2	T30289	pristinamycin I sy
41	29	70.7	118	2	S40374	Ig kappa chain - h
42	29	70.7	118	2	S52262	histone H4 - Entam
43	29	70.7	167	2	D71914	nonheme iron-conta
44	29	70.7	168	2	T22949	hypothetical prote
45	29	70.7	217	2	F82612	hypothetical prote

ALIGNMENTS

RESULT 1
S39358
cyclin kinase inhibitor - human (fragments)
C;Species: Homo sapiens (man)
C;Date: 25-Feb-1994 #sequence_revision 17-Nov-1995 #text_change 17-Mar-1999
C;Accession: S39358
R;Xiong, Y.; Hannon, G.J.; Zhang, H.; Casso, D.; Kobayashi, R.; Beach, D.
Nature 366, 701-704, 1993
A;Title: p21 is a universal inhibitor of cyclin kinases.
A;Reference number: S39357; MUID:94081955; PMID:8259214
A;Accession: S39358
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-47 <XIO>

Query Match 92.7%; Score 38; DB 2; Length 47;
Best Local Similarity 87.5%; Pred. No. 0.28;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 HAKRRLIF 8
             |:||||||
Db      38 HSKRRLIF 45
```

RESULT 2
I54380
cyclin-dependent kinase - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C;Accession: I54380
R;Mousses, S.; Ozcelik, H.; Lee, P.D.; Malkin, D.; Bull, S.B.; Andrulis, I.L.
Hum. Mol. Genet. 4, 1089-1092, 1995
A;Title: Two variants of the CIP1/WAF1 gene occur together and are associated with hu
A;Reference number: I54380; MUID:95384154; PMID:7655464
A;Accession: I54380
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-181 <RES>
A;Cross-references: GB:L47232; NID:g984723; PIDN:AAB59559.1; PID:g984724
C;Genetics:
A;Gene: CIP1/WAF1

Query Match 92.7%; Score 38; DB 2; Length 181;
Best Local Similarity 87.5%; Pred. No. 0.98;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 HAKRRLIF 8
             |:||||||
Db     169 HSKRRLIF 176
```

RESULT 3

I68674
cyclin-dependent kinase - human (fragment)
N;Alternate names: probable DNA synthesis inhibitor
C;Species: Homo sapiens (man)
C;Date: 04-Oct-1996 #sequence_revision 04-Oct-1996 #text_change 01-Dec-2000
C;Accession: I68674; A49437; I53412; S39357
R;Mousses, S.; Ozcelik, H.; Lee, P.D.; Malkin, D.; Bull, S.B.; Andrulis, I.L.
Hum. Mol. Genet. 4, 1089-1092, 1995
A;Title: Two variants of the CIP1/WAF1 gene occur together and are associated with human
A;Reference number: I54380; MUID:95384154; PMID:7655464
A;Accession: I68674
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-181 <RES>
A;Cross-references: GB:L47233; NID:g986878; PIDN:AAB59560.1; PID:g986879
R;Harper, J.W.; Adami, G.R.; Wei, N.; Keyomarsi, K.; Elledge, S.J.
Cell 75, 805-816, 1993
A;Title: The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-depende
A;Reference number: A49437; MUID:94061996; PMID:8242751
A;Accession: A49437
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 18-181 <RE3>
A;Cross-references: GB:L25610; NID:g425142; PIDN:AAA16109.1; PID:g425143
R;Noda, A.; Ning, Y.; Venable, S.F.; Pereira-Smith, O.M.; Smith, J.R.
Exp. Cell Res. 211, 90-98, 1994
A;Title: Cloning of senescent cell-derived inhibitors of DNA synthesis using an expressi
A;Reference number: I53412; MUID:94170884; PMID:8125163
A;Accession: I53412
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 18-181 <RE2>
A;Cross-references: GB:L26165; NID:g418017; PIDN:AAA19811.1; PID:g433742
R;Xiong, Y.; Hannon, G.J.; Zhang, H.; Casso, D.; Kobayashi, R.; Beach, D.
Nature 366, 701-704, 1993
A;Title: p21 is a universal inhibitor of cyclin kinases.
A;Reference number: S39357; MUID:94081955; PMID:8259214
A;Accession: S39357
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 18-181 <XIO>
A;Cross-references: GB:S67388; NID:g453134; PIDN:AAB29246.1; PID:g453135
C;Genetics:
A;Gene: CIP1/WAF1

  Query Match 92.7%; Score 38; DB 2; Length 181;
  Best Local Similarity 87.5%; Pred. No. 0.98;
  Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HAKRRLIF 8
|:|||||
Db 169 HSKRRLIF 176

RESULT 4
I49023
tumor suppressor p21 WAF1/CIP1 [imported] - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Jun-2000
C;Accession: I49023; I49296
R;Huppi, K.; Siwarski, D.; Dosik, J.; Michieli, P.; Chedid, M.; Reed, S.; Mock, B.; Givo
Oncogene 9, 3017-3020, 1994
A;Title: Molecular cloning, sequencing, chromosomal localization and expression of mouse
A;Reference number: I49023; MUID:94366751; PMID:8084607
A;Accession: I49023
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-159 <RES>
A;Cross-references: EMBL:U09507; NID:g595302; PIDN:AAB60456.1; PID:g595303
R;El-Deiry, W.S.; Tokino, T.; Waldman, T.; Velculescu, V.; Oliner, J.D.; Burell, M.; Hil
Cancer Res. 55, 2910-2919, 1995
A;Title: Topological control of p21WAF1/CIP1 expression in normal and neoplastic tissues
A;Reference number: I49296; MUID:95316868; PMID:7796420
A;Accession: I49296
A;Status: nucleic acid sequence not shown; translation not shown; translated from GB/
A;Molecule type: mRNA
A;Residues: 1-159 <RE2>
A;Cross-references: EMBL:U24173; NID:g902578; PIDN:AAC52220.1; PID:g902579
C;Genetics:
A;Gene: Waf1

  Query Match 90.2%; Score 37; DB 2; Length 159;
  Best Local Similarity 75.0%; Pred. No. 1.4;
  Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HAKRRLIF 8
|:||||:|
Db 147 HSKRRLVF 154

RESULT 5
I84725
tumor suppressor p21 WAF1/CIP1 [imported] - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 20-Jun-2000
C;Accession: I84725
R;El-Deiry, W.S.; Tokino, T.; Waldman, T.; Velculescu, V.; Oliner, J.D.; Burell, M.;
Cancer Res. 55, 2910-2919, 1995
A;Title: Topological control of p21WAF1/CIP1 expression in normal and neoplastic tiss
A;Reference number: I49296; MUID:95316868; PMID:7796420
A;Accession: I84725
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-164 <RES>
A;Cross-references: EMBL:U24174; NID:g902581; PIDN:AAC52221.1; PID:g902582
C;Genetics:
A;Gene: WAF1

  Query Match 90.2%; Score 37; DB 2; Length 164;
  Best Local Similarity 75.0%; Pred. No. 1.5;
  Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HAKRRLIF 8
|:||||:|
Db 152 HSKRRLVF 159

RESULT 6
S41986
26S proteasome regulatory particle chain RPN4 - yeast (Saccharomyces cerevisiae)
N;Alternate names: nuclear protein SON1; protein D2840; protein L00928; protein YDL02
C;Species: Saccharomyces cerevisiae
C;Date: 31-Mar-1992 #sequence_revision 14-Sep-1994 #text_change 29-Oct-1999
C;Accession: S41986; S52499; S67552; S30806
R;Nelson, M.K.; Kurihara, T.; Silver, P.A.
Genetics 134, 159-173, 1993
A;Title: Extragenic suppressors of mutations in the cytoplasmic C terminus of SEC63 d
A;Reference number: S41986; MUID:93292918; PMID:8514125
A;Accession: S41986
A;Molecule type: DNA
A;Residues: 1-531 <NEL>
A;Cross-references: EMBL:L00928; NID:g172650; PIDN:AAA35067.1; PID:g172651
R;Andre, B.; Vissers, S.; Urrestarazu, L.
submitted to the EMBL Data Library, February 1995
A;Description: The sequence of a 42 kb segment located on the left arm of chromosome
A;Reference number: S52492
A;Accession: S52499
A;Molecule type: DNA
A;Residues: 1-531 <AND>
A;Cross-references: EMBL:Z48432; NID:g683669; PIDN:CAA88339.1; PID:g683677
R;Urrestarazu, L.A.; Andre, B.; Vissers, S.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67535
A;Accession: S67552
A;Molecule type: DNA
A;Residues: 1-531 <URR>

```
A;Cross-references: EMBL:Z74068; NID:g1430989; PIDN:CAA98579.1; PID:e252979; PID:g143099
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:RPN4; SON1; UFD5
A;Cross-references: SGD:S0002178; MIPS:YDL020c
A;Map position: 4L
C;Keywords: nucleus
F;211-229/Region: acidic
F;300-312/Region: acidic

  Query Match             85.4%;  Score 35;  DB 2;  Length 531;
  Best Local Similarity   62.5%;  Pred. No. 12;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          ||||:::|
Db    468 HAKRKIVF 475


RESULT 7
H89959
conserved hypothetical protein SA1569 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 01-Feb-2002
C;Accession: H89959
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: H89959
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-214 <KUR>
A;Cross-references: GB:BA000018; PID:g13701543; PIDN:BAB42837.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA1569
C;Superfamily: hypothetical protein HI0340

  Query Match             80.5%;  Score 33;  DB 2;  Length 214;
  Best Local Similarity   75.0%;  Pred. No. 14;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          ||||||  :
Db    124 HAKRRLTY 131


RESULT 8
E96619
protein T30E16.11 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: E96619
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: E96619
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-76 <STO>
A;Cross-references: GB:AE005173; NID:g8778740; PIDN:AAF79748.1; GSPDB:GN00141
C;Genetics:
A;Gene: T30E16.11
A;Map position: 1

  Query Match             75.6%;  Score 31;  DB 2;  Length 76;
  Best Local Similarity   62.5%;  Pred. No. 14;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
            | |||  :|
Db     47 HGKRRFLF 54


RESULT 9
AG2036
hypothetical protein alr1845 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AG2036
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG2036
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-219 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB73544.1; PID:g17130935; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr1845
C;Superfamily: hypothetical protein HI0340

  Query Match             75.6%;  Score 31;  DB 2;  Length 219;
  Best Local Similarity   71.4%;  Pred. No. 38;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLI 7
          |||||::
Db    132 HAKRRVV 138


RESULT 10
S60545
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-45-1) (fr
C;Species: human immunodeficiency virus type 1, HIV-1
A;Variety: isolate CI-45-1
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C;Accession: S60545
R;Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J
AIDS 8, 21-26, 1994
A;Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cot
A;Reference number: S60521; MUID:94280700; PMID:8011235
A;Accession: S60545
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-294 <JAN>
A;Cross-references: EMBL:X72047; NID:g468669; PIDN:CAA50930.1; PID:g468670
A;Experimental source: isolate CI-45-1
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Genetics:
A;Gene: env
C;Superfamily: type E retrovirus env polyprotein
C;Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

  Query Match             75.6%;  Score 31;  DB 2;  Length 294;
  Best Local Similarity   71.4%;  Pred. No. 50;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLI 7
          |||||::
Db    225 HAKRRVV 231
```

RESULT 11
S60524
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-45-3) (fragm
C;Species: human immunodeficiency virus type 1, HIV-1
A;Variety: isolate CI-45-3
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C;Accession: S60524
R;Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J.;
AIDS 8, 21-26, 1994
A;Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cote d
A;Reference number: S60521; MUID:94280700; PMID:8011235
A;Accession: S60524
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-294 <JAN>
A;Cross-references: EMBL:X72027; NID:g468780; PIDN:CAA50910.1; PID:g468781
A;Experimental source: isolate CI-45-3
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Genetics:
A;Gene: env
C;Superfamily: type E retrovirus env polyprotein
C;Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

  Query Match 75.6%; Score 31; DB 2; Length 294;
  Best Local Similarity 71.4%; Pred. No. 50;
  Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HAKRRLI 7
|||||::
Db 225 HAKRRVV 231

RESULT 12
A38705
heterocyst development protein hetR - Anabaena sp. (strain PCC 7120)
C;Species: Anabaena sp.
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: A38705
R;Buikema, W.J.; Haselkorn, R.
Genes Dev. 5, 321-330, 1991
A;Title: Characterization of a gene controlling heterocyst differentiation in the cyanob
A;Reference number: A38705; MUID:91138965; PMID:1840555
A;Accession: A38705
A;Molecule type: DNA
A;Residues: 1-299 <BUI>
A;Cross-references: GB:M37779; NID:g142021; PIDN:AAA21998.1; PID:g142022
C;Comment: This protein is required for and probably controls heterocyst development.
C;Superfamily: Anabaena heterocyst development protein hetR

  Query Match 75.6%; Score 31; DB 1; Length 299;
  Best Local Similarity 62.5%; Pred. No. 50;
  Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HAKRRLIF 8
| ||||::
Db 185 HIKRRLLY 192

RESULT 13
S60529
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-14-13) (frag
C;Species: human immunodeficiency virus type 1, HIV-1
A;Variety: isolate CI-14-13
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C;Accession: S60529
R;Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J.;
AIDS 8, 21-26, 1994
A;Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cote d
A;Reference number: S60521; MUID:94280700; PMID:8011235
A;Accession: S60529
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-299 <JAN>
A;Cross-references: EMBL:X72031; NID:g468637; PIDN:CAA50914.1; PID:g468638
A;Experimental source: isolate CI-14-13
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Genetics:
A;Gene: env
C;Superfamily: type E retrovirus env polyprotein
C;Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

  Query Match 75.6%; Score 31; DB 2; Length 299;
  Best Local Similarity 71.4%; Pred. No. 50;
  Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HAKRRLI 7
|||||::
Db 230 HAKRRVV 236

RESULT 14
S60552
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-327-2) (f
C;Species: human immunodeficiency virus type 1, HIV-1
A;Variety: isolate CI-327-2
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C;Accession: S60552
R;Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J
AIDS 8, 21-26, 1994
A;Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cot
A;Reference number: S60521; MUID:94280700; PMID:8011235
A;Accession: S60552
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-299 <JAN>
A;Cross-references: EMBL:X72056; NID:g468685; PIDN:CAA50937.1; PID:g468686
A;Experimental source: isolate CI-327-2
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Genetics:
A;Gene: env
C;Superfamily: type E retrovirus env polyprotein
C;Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

  Query Match 75.6%; Score 31; DB 2; Length 299;
  Best Local Similarity 71.4%; Pred. No. 50;
  Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HAKRRLI 7
|||||::
Db 230 HAKRRVV 236

RESULT 15
S60553
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-327-3) (f
C;Species: human immunodeficiency virus type 1, HIV-1
A;Variety: isolate CI-327-3
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C;Accession: S60553
R;Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J
AIDS 8, 21-26, 1994
A;Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cot
A;Reference number: S60521; MUID:94280700; PMID:8011235
A;Accession: S60553
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-299 <JAN>
A;Cross-references: EMBL:X72057; NID:g468687; PIDN:CAA50938.1; PID:g468688
A;Experimental source: isolate CI-327-3
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C;Genetics:
A;Gene: env
C;Superfamily: type E retrovirus env polyprotein
C;Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match 75.6%; Score 31; DB 2; Length 299;
Best Local Similarity 71.4%; Pred. No. 50;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 HAKRRLI 7
             |||||::
Db     230 HAKRRVV 236
```

Search completed: December 14, 2002, 15:50:07
Job time : 31.5 secs

Functional frags of SEQ A

~~then~~ & SEQ is Enabled

& functional, then why

Doctrine of Equivalents

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: December 14, 2002, 15:41:54 ; Search time 57.5 Seconds
(without alignments)
28.667 Million cell updates/sec

Title: US-09-726-470A-35
Perfect score: 41
Sequence: 1 HAKRRLIF 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	92.7	181	4	Q14010	Q14010 homo sapien
2	37	90.2	164	11	Q64315	Q64315 rattus norv
3	33	80.5	149	2	O30605	O30605 trichodesmi
4	33	80.5	149	2	O30606	O30606 trichodesmi
5	33	80.5	149	2	O30607	O30607 trichodesmi
6	33	80.5	149	2	O30608	O30608 leptolyngby
7	33	80.5	149	2	Q9ZFZ6	Q9zfz6 trichodesmi
8	33	80.5	149	2	Q9ZFZ5	Q9zfz5 trichodesmi
9	33	80.5	149	2	Q9RQN4	Q9rqn4 trichodesmi
10	33	80.5	214	16	Q99TB6	Q99tb6 staphylococ
11	33	80.5	302	2	Q93CE9	Q93ce9 trichodesmi
12	33	80.5	365	10	Q8RYI8	Q8ryi8 oryza sativ
13	33	80.5	580	16	Q9K4J4	Q9k4j4 streptomyce
14	33	80.5	721	10	Q8W3G0	Q8w3g0 oryza sativ
15	32	78.0	288	5	O96820	O96820 plasmodium
16	31	75.6	76	10	Q9LQ60	Q9lq60 arabidopsis
17	31	75.6	100	8	Q9MDV6	Q9mdv6 beta vulgar
18	31	75.6	149	2	Q99QB2	Q99qb2 nodularia h
19	31	75.6	149	2	Q99Q50	Q99q50 nodularia s
20	31	75.6	149	2	O30609	O30609 symploca sp
21	31	75.6	149	2	Q9XCP1	Q9xcp1 fischerella
22	31	75.6	149	2	Q9XCP0	Q9xcp0 nodularia s
23	31	75.6	149	2	Q9RAH5	Q9rah5 nostoc sp.
24	31	75.6	149	2	Q9XCN9	Q9xcn9 richelia sp
25	31	75.6	149	2	Q9RAC9	Q9rac9 richelia sp
26	31	75.6	149	2	Q9XCN8	Q9xcn8 richelia sp
27	31	75.6	149	2	Q9RAC8	Q9rac8 richelia sp
28	31	75.6	149	2	Q9RAC7	Q9rac7 richelia sp
29	31	75.6	149	2	Q9RAC6	Q9rac6 richelia sp
30	31	75.6	149	2	Q9RAC5	Q9rac5 richelia sp
31	31	75.6	149	2	Q9RAC4	Q9rac4 richelia sp
32	31	75.6	149	2	Q9L8X5	Q9l8x5 anabaena ge
33	31	75.6	149	2	Q9L8X4	Q9l8x4 aphanizomen
34	31	75.6	149	2	Q9L8X3	Q9l8x3 anabaena sp
35	31	75.6	149	2	Q9AP95	Q9ap95 anabaena sp
36	31	75.6	149	2	Q9AP94	Q9ap94 anabaena sp
37	31	75.6	149	2	Q9AP93	Q9ap93 anabaena sp
38	31	75.6	149	2	Q9ANW3	Q9anw3 nodularia s
39	31	75.6	149	2	Q9XCP3	Q9xcp3 calothrix s
40	31	75.6	149	2	Q9XCP2	Q9xcp2 calothrix s
41	31	75.6	219	16	Q8YVX4	Q8yvx4 anabaena sp
42	31	75.6	294	15	Q76168	Q76168 human immun
43	31	75.6	294	15	Q76190	Q76190 human immun
44	31	75.6	298	2	Q51320	Q51320 nostoc sp.
45	31	75.6	299	2	Q9AH77	Q9ah77 nostoc punc

ALIGNMENTS

RESULT 1
Q14010
ID Q14010 PRELIMINARY; PRT; 181 AA.
AC Q14010;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Cyclin-dependent kinase (Fragment).
GN CIP1/WAF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TUMOR;
RX MEDLINE=95384154; PubMed=7655464;
RA Mousses S., Ozcelik H., Lee P.D., Malkin D., Bull S.B., Andrulis I.L.;
RT "Two variants of the CIP1/WAF1 gene occur together and are associated
RT with human cancer.";
RL Hum. Mol. Genet. 4:1089-1092(1995).
DR EMBL; L47232; AAB59559.1; -.
DR InterPro; IPR003175; CDI.
DR Pfam; PF02234; CDI; 1.
KW Kinase.
FT NON_TER 1 1
SQ SEQUENCE 181 AA; 20083 MW; 4CCFA5112232D4F1 CRC64;

Query Match 92.7%; Score 38; DB 4; Length 181;
Best Local Similarity 87.5%; Pred. No. 1.8;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 HAKRRLIF 8
             |:||||| 
Db     169 HSKRRLIF 176
```

RESULT 2
Q64315

```
ID   Q64315      PRELIMINARY;      PRT;   164 AA.
AC   Q64315;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   P21 (WAF1).
GN   WAF1 OR CIP1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95316868; PubMed=7796420;
RA   el-Deiry W.S., Tokino T., Waldman T., Velculescu V., Oliner J.D.,
RA   Burell M., Hill D.E., Rees J.L., Hamilton S.R., Kinzler K.W.,
RA   Vogelstein B.;
RT   "Topological control of p21WAF1/CIP1 expression in normal and
RT   neoplastic tissues.";
RL   Cancer Res. 55:2910-2919(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F344/N; TISSUE=LUNG;
RA   Belinsky S.A.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U24174; AAC52221.1; -.
DR   EMBL; L41275; AAC42084.1; -.
DR   InterPro; IPR003175; CDI.
DR   Pfam; PF02234; CDI; 1.
SQ   SEQUENCE   164 AA;  18318 MW;  6057E86045B6435F CRC64;

  Query Match             90.2%;  Score 37;  DB 11;  Length 164;
  Best Local Similarity   75.0%;  Pred. No. 2.7;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          |:|||||:|
Db    152 HSKRRLVF 159


RESULT 3
O30605
ID   O30605      PRELIMINARY;      PRT;   149 AA.
AC   O30605;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Heterocyst differentiation protein (Fragment).
GN   HETR.
OS   Trichodesmium contortum.
OC   Bacteria; Cyanobacteria; Oscillatoriales; Trichodesmium.
OX   NCBI_TaxID=64179;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Janson S., Bergman B., Carpenter E.J., Giovannoni S.J., Vergin K.;
RT   "Genetic analysis ofnatural populations of the marine diazotrophic
RT   cyanobacterium Trichodesmium.";
RL   FEMS Microbiol. Ecol. 30:57-65(1999).
DR   EMBL; AF013031; AAB81941.1; -.
DR   MEROPS; S48.001; -.
DR   InterPro; IPR005319; Peptidase_S48.
DR   Pfam; PF03574; Peptidase_S48; 1.
FT   NON_TER       1      1
FT   NON_TER     149    149
SQ   SEQUENCE   149 AA;  17610 MW;  9DC13448BEAD83CC CRC64;

  Query Match             80.5%;  Score 33;  DB 2;  Length 149;
  Best Local Similarity   75.0%;  Pred. No. 19;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          | |||||:
Db    132 HIKRRLIY 139


RESULT 4
O30606
ID   O30606      PRELIMINARY;      PRT;   149 AA.
AC   O30606;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Heterocyst differentiation protein (Fragment).
GN   HETR.
OS   Trichodesmium hildebrandtii.
OC   Bacteria; Cyanobacteria; Oscillatoriales; Trichodesmium.
OX   NCBI_TaxID=84114;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Janson S., Bergman B., Carpenter E.J., Giovannoni S.J., Vergin K.;
RT   "Genetic analysis ofnatural populations of the marine diazotrophic
RT   cyanobacterium Trichodesmium.";
RL   FEMS Microbiol. Ecol. 30:57-65(1999).
DR   EMBL; AF013032; AAC90368.1; -.
DR   MEROPS; S48.001; -.
DR   InterPro; IPR005319; Peptidase_S48.
DR   Pfam; PF03574; Peptidase_S48; 1.
FT   NON_TER       1      1
FT   NON_TER     149    149
SQ   SEQUENCE   149 AA;  17364 MW;  376AAA2F42F56C66 CRC64;

  Query Match             80.5%;  Score 33;  DB 2;  Length 149;
  Best Local Similarity   75.0%;  Pred. No. 19;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          | |||||:
Db    132 HIKRRLIY 139


RESULT 5
O30607
ID   O30607      PRELIMINARY;      PRT;   149 AA.
AC   O30607;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Heterocyst differentiation protein (Fragment).
GN   HETR.
OS   Trichodesmium tenue.
OC   Bacteria; Cyanobacteria; Oscillatoriales; Trichodesmium.
OX   NCBI_TaxID=64180;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Janson S., Bergman B., Carpenter E.J., Giovannoni S.J., Vergin K.;
RT   "Genetic analysis ofnatural populations of the marine diazotrophic
RT   cyanobacterium Trichodesmium.";
RL   FEMS Microbiol. Ecol. 30:57-65(1999).
DR   EMBL; AF013033; AAB81942.1; -.
DR   MEROPS; S48.001; -.
DR   InterPro; IPR005319; Peptidase_S48.
DR   Pfam; PF03574; Peptidase_S48; 1.
FT   NON_TER       1      1
FT   NON_TER     149    149
SQ   SEQUENCE   149 AA;  17603 MW;  8156A990DAAFF4E4 CRC64;

  Query Match             80.5%;  Score 33;  DB 2;  Length 149;
  Best Local Similarity   75.0%;  Pred. No. 19;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          | |||||:
Db    132 HIKRRLIY 139


RESULT 6
```

```
O30608
ID   O30608      PRELIMINARY;      PRT;   149 AA.
AC   O30608;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Heterocyst differentiation protein (Fragment).
GN   HETR.
OS   Leptolyngbya sp. PCC 73110.
OC   Bacteria; Cyanobacteria; Oscillatoriales; Leptolyngbya.
OX   NCBI_TaxID=102128;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC73110;
RX   MEDLINE=99052091; PubMed=9835026;
RA   Janson S., Matveyev A., Bergman B.;
RT   "The presence and expression of hetR in the non-heterocystous
RT   cyanobacterium Symploca PCC 8002.";
RL   FEMS Microbiol. Lett. 168:173-179(1998).
DR   EMBL; AF013034; AAC34930.1; -.
DR   MEROPS; S48.001; -.
DR   InterPro; IPR005319; Peptidase_S48.
DR   Pfam; PF03574; Peptidase_S48; 1.
FT   NON_TER       1      1
FT   NON_TER     149    149
SQ   SEQUENCE   149 AA;  17628 MW;  328DD4C14D7A32A6 CRC64;

  Query Match             80.5%;  Score 33;  DB 2;  Length 149;
  Best Local Similarity   75.0%;  Pred. No. 19;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          | |||||:
Db    132 HIKRRLIY 139


RESULT 7
Q9ZFZ6
ID   Q9ZFZ6      PRELIMINARY;      PRT;   149 AA.
AC   Q9ZFZ6;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Heterocyst differentiation protein (Fragment).
GN   HETR.
OS   Trichodesmium sp. (strain IMS101).
OC   Bacteria; Cyanobacteria; Oscillatoriales; Trichodesmium.
OX   NCBI_TaxID=57878;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IMS 101;
RA   Janson S., Bergman B., Carpenter E.J., Giovannoni S.J., Vergin K.;
RT   "Genetic analysis ofnatural populations of the marine diazotrophic
RT   cyanobacterium Trichodesmium.";
RL   FEMS Microbiol. Ecol. 30:57-65(1999).
DR   EMBL; AF091323; AAC95053.1; -.
DR   MEROPS; S48.001; -.
DR   InterPro; IPR005319; Peptidase_S48.
DR   Pfam; PF03574; Peptidase_S48; 1.
FT   NON_TER       1      1
FT   NON_TER     149    149
SQ   SEQUENCE   149 AA;  17628 MW;  328DD4C14D7A32A6 CRC64;

  Query Match             80.5%;  Score 33;  DB 2;  Length 149;
  Best Local Similarity   75.0%;  Pred. No. 19;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          | |||||:
Db    132 HIKRRLIY 139


RESULT 8
Q9ZFZ5
ID   Q9ZFZ5      PRELIMINARY;      PRT;   149 AA.
AC   Q9ZFZ5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Heterocyst differentiation protein (Fragment).
GN   HETR.
OS   Trichodesmium thiebautii.
OC   Bacteria; Cyanobacteria; Oscillatoriales; Trichodesmium.
OX   NCBI_TaxID=1208;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PUFF;
RA   Janson S., Bergman B., Carpenter E.J., Giovannoni S.J., Vergin K.;
RT   "Genetic analysis ofnatural populations of the marine diazotrophic
RT   cyanobacterium Trichodesmium.";
RL   FEMS Microbiol. Ecol. 30:57-65(1999).
DR   EMBL; AF091324; AAC95054.1; -.
DR   MEROPS; S48.001; -.
DR   InterPro; IPR005319; Peptidase_S48.
DR   Pfam; PF03574; Peptidase_S48; 1.
FT   NON_TER       1      1
FT   NON_TER     149    149
SQ   SEQUENCE   149 AA;  17188 MW;  05FBC643DC635EBC CRC64;

  Query Match             80.5%;  Score 33;  DB 2;  Length 149;
  Best Local Similarity   75.0%;  Pred. No. 19;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          | |||||:
Db    132 HIKRRLIY 139


RESULT 9
Q9RQN4
ID   Q9RQN4      PRELIMINARY;      PRT;   149 AA.
AC   Q9RQN4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Heterocyst differentiation protein (Fragment).
GN   HETR.
OS   Trichodesmium thiebautii.
OC   Bacteria; Cyanobacteria; Oscillatoriales; Trichodesmium.
OX   NCBI_TaxID=1208;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Janson S., Bergman B., Carpenter E.J., Giovannoni S.J., Vergin K.;
RT   "Genetic analysis ofnatural populations of the marine diazotrophic
RT   cyanobacterium Trichodesmium.";
RL   FEMS Microbiol. Ecol. 30:57-65(1999).
DR   EMBL; AF091325; AAC95055.1; -.
DR   MEROPS; S48.001; -.
DR   InterPro; IPR005319; Peptidase_S48.
DR   Pfam; PF03574; Peptidase_S48; 1.
FT   NON_TER       1      1
FT   NON_TER     149    149
SQ   SEQUENCE   149 AA;  17327 MW;  2A30DAEA3E5C608C CRC64;

  Query Match             80.5%;  Score 33;  DB 2;  Length 149;
  Best Local Similarity   75.0%;  Pred. No. 19;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          | |||||:
Db    132 HIKRRLIY 139


RESULT 10
Q99TB6
ID   Q99TB6      PRELIMINARY;      PRT;   214 AA.
```

```
AC   Q99TB6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein SAV1748.
GN   SAV1748 OR SA1569.
OS   Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS   Staphylococcus aureus (strain N315).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Staphylococcus.
OX   NCBI_TaxID=158878, 158879;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX   MEDLINE=21311952; PubMed=11418146;
RA   Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA   Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA   Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA   Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA   Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA   Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA   Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT   "Whole genome sequencing of meticillin-resistant Staphylococcus
RT   aureus.";
RL   Lancet 357:1225-1240(2001).
DR   EMBL; AP003363; BAB57910.1; -.
DR   EMBL; AP003134; BAB42837.1; -.
DR   InterPro; IPR004395; Cons_hypoth91.
DR   InterPro; IPR003358; Methyltransf_4.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF02390; Methyltransf_4; 1.
DR   TIGRFAMs; TIGR00091; Cons_hypoth91; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   214 AA;  25291 MW;  6E0EA5A49A3FF264 CRC64;

  Query Match             80.5%;  Score 33;  DB 16;  Length 214;
  Best Local Similarity   75.0%;  Pred. No. 26;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          |||||| :
Db    124 HAKRRLTY 131


RESULT 11
Q93CE9
ID   Q93CE9      PRELIMINARY;      PRT;   302 AA.
AC   Q93CE9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Heterocyst differentiation protein.
GN   HETR.
OS   Trichodesmium sp. (strain IMS101).
OC   Bacteria; Cyanobacteria; Oscillatoriales; Trichodesmium.
OX   NCBI_TaxID=57878;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IMS101;
RA   Schiefer W.;
RT   "Characterization of the hetR gene in the cyanobacterium
RT   Trichodesmium.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF410432; AAL05045.1; -.
DR   InterPro; IPR005319; Peptidase_S48.
DR   Pfam; PF03574; Peptidase_S48; 1.
SQ   SEQUENCE   302 AA;  35438 MW;  784FC4FAACBC4A68 CRC64;

  Query Match             80.5%;  Score 33;  DB 2;  Length 302;
  Best Local Similarity   75.0%;  Pred. No. 36;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          | ||||:
Db    186 HIKRRLIY 193


RESULT 12
Q8RYI8
ID   Q8RYI8      PRELIMINARY;      PRT;   365 AA.
AC   Q8RYI8;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   OSJNBa0042P21.10 protein.
GN   OSJNBA0042P21.10.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC
RT   clone:OSJNBa0042P21.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP004614; BAB90802.1; -.
SQ   SEQUENCE   365 AA;  41228 MW;  5E968975DB77E3E7 CRC64;

  Query Match             80.5%;  Score 33;  DB 10;  Length 365;
  Best Local Similarity   85.7%;  Pred. No. 43;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLI 7
          |||||:|
Db    236 HAKRRII 242


RESULT 13
Q9K4J4
ID   Q9K4J4      PRELIMINARY;      PRT;   580 AA.
AC   Q9K4J4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein SCO7320.
GN   SCO7320 OR SC5F8.30C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
```

```
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL357613; CAB93758.1; -.
DR   InterPro; IPR003018; GAF.
DR   InterPro; IPR001932; PP2C-like.
DR   SMART; SM00065; GAF; 1.
DR   SMART; SM00331; PP2C_SIG; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   580 AA;  61543 MW;  BC7E7EC2332DB051 CRC64;

  Query Match             80.5%;  Score 33;  DB 16;  Length 580;
  Best Local Similarity   75.0%;  Pred. No. 65;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 HAKRRLIF 8
           ||:||| |
Db     191 HARRRLTF 198


RESULT 14
Q8W3G0
ID   Q8W3G0      PRELIMINARY;      PRT;   721 AA.
AC   Q8W3G0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative far-red impaired response protein.
GN   OSJNBA0035H01.9.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA   Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA   Tsitrin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pai G.,
RA   VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA   Salzberg S.L., White O., Fraser C.M.;
RT   "Oryza sativa chromosome 10 BAC OSJNBa0035H01 genomic sequence.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC027037; AAL58182.1; -.
DR   InterPro; IPR004330; FAR1.
DR   InterPro; IPR003653; SUMO_protease.
DR   Pfam; PF03101; FAR1; 1.
DR   Pfam; PF02902; Peptidase_C48; 1.
SQ   SEQUENCE   721 AA;  83859 MW;  FA4B92205458BC4E CRC64;

  Query Match             80.5%;  Score 33;  DB 10;  Length 721;
  Best Local Similarity   85.7%;  Pred. No. 80;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 HAKRRLI 7
           |||||:|
Db     611 HAKRRII 617


RESULT 15
O96820
ID   O96820      PRELIMINARY;      PRT;   288 AA.
AC   O96820;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Cdc2-related kinase 2.
OS   Plasmodium berghei (strain Anka).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANKA;
RX   MEDLINE=99020256; PubMed=9803415;
RA   Vinkenoog R., Speranca M.A., Ramesar J., Thomas A.W.,
RA   del Portillo H.A., Janse C.J., Waters A.P.;
RT   "Characterisation of the Cdc2-related kinase 2 gene from Plasmodium
RT   knowlesi and P. berghei.";
RL   Mol. Biochem. Parasitol. 95:229-240(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANKA;
RX   MEDLINE=96065755; PubMed=7477105;
RA   Vinkenoog R., Veldhuisen B., Speranca M.A., del Portillo H.A.,
RA   Janse C.J., Waters A.P.;
RT   "Comparison of introns in a cdc2-homologous gene in a number of
RT   Plasmodium species.";
RL   Mol. Biochem. Parasitol. 71:233-241(1995).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AJ224152; CAA11849.1; -.
DR   HSSP; P24941; 1HCL.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   288 AA;  33019 MW;  EFB28585AB2AF124 CRC64;

  Query Match             78.0%;  Score 32;  DB 5;  Length 288;
  Best Local Similarity   85.7%;  Pred. No. 57;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 HAKRRLI 7
           |||:|||
Db      70 HAKKRLI 76


Search completed: December 14, 2002, 15:48:54
Job time : 59.5 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: December 14, 2002, 15:34:14 ; Search time 23.5 Seconds
(without alignments)
14.120 Million cell updates/sec

Title: US-09-726-470A-35
Perfect score: 41
Sequence: 1 HAKRRLIF 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	92.7	164	1	CDN1_FELCA	O19002 felis silve
2	38	92.7	164	1	CDN1_HUMAN	P38936 homo sapien
3	37	90.2	159	1	CDN1_MOUSE	P39689 mus musculu
4	35	85.4	531	1	RPN4_YEAST	Q03465 saccharomyc
5	33	80.5	214	1	MT04_STAAU	Q9kwz4 staphylococ
6	31	75.6	202	1	MT04_MYCHO	Q9f411 mycoplasma
7	31	75.6	299	1	HETR_ANASP	P27709 anabaena sp
8	31	75.6	487	1	PHOQ_SALTY	P14147 salmonella
9	30	73.2	141	1	YEF5_YEAST	P32616 saccharomyc
10	30	73.2	250	1	MT04_TREPA	O83477 treponema p
11	30	73.2	751	1	YK09_YEAST	P36124 saccharomyc
12	29	70.7	118	1	H4_ENTHI	P40287 entamoeba h
13	29	70.7	167	1	FTN_HELPJ	Q9zlil helicobacte
14	29	70.7	206	1	TPIS_AEDTO	P92119 aedes togoi
15	29	70.7	206	1	TPIS_ANOME	P91895 anopheles m
16	29	70.7	206	1	TPIS_CULPI	P91919 culex pipie
17	29	70.7	215	1	TPIS_HELVI	P55275 heliothis v
18	29	70.7	238	1	MT04_NEIMA	Q9ju19 neisseria m
19	29	70.7	238	1	MT04_NEIMB	Q9jz24 neisseria m
20	29	70.7	244	1	MT04_XYLFA	Q9pf94 xylella fas
21	29	70.7	246	1	TPIS_CULTA	P30741 culex tarsa
22	29	70.7	247	1	TPIS_DROME	P29613 drosophila
23	29	70.7	248	1	TPIS_SCHPO	P07669 schizosacch
24	29	70.7	252	1	YRY2_CAEEL	Q10006 caenorhabdi
25	29	70.7	271	1	MT04_CAUCR	P58088 caulobacter
26	29	70.7	271	1	MT04_STRCO	Q9f305 streptomyce
27	29	70.7	327	1	RL5_ANOGA	O44248 anopheles g
28	29	70.7	831	1	RECG_SYNY3	Q55681 synechocyst
29	29	70.7	855	1	GCFC_MOUSE	P58501 mus musculu
30	29	70.7	917	1	GCFC_HUMAN	Q9y5b6 homo sapien
31	29	70.7	959	1	G2D1_HUMAN	Q9uhl9 h general t
32	29	70.7	1104	1	G2D1_MOUSE	Q9ji57 mus musculu
33	28	68.3	40	1	PSAI_PEA	P17227 pisum sativ
34	28	68.3	167	1	FTN_HELPY	P52093 helicobacte
35	28	68.3	168	1	OLE6_GOSHI	P29527 gossypium h
36	28	68.3	192	1	RL24_SCHPO	Q10353 schizosacch
37	28	68.3	211	1	MT04_STRPY	P58090 streptococc
38	28	68.3	213	1	MT04_BACSU	O34522 bacillus su
39	28	68.3	217	1	MT04_LACLA	Q9chi2 lactococcus
40	28	68.3	220	1	MT04_BACHD	Q9k7u8 bacillus ha
41	28	68.3	224	1	MT04_UREPA	Q9pqm2 ureaplasma
42	28	68.3	251	1	TPIS_PSEAE	Q9hv51 pseudomonas
43	28	68.3	252	1	TPIS_SCHJA	Q27775 schistosoma
44	28	68.3	261	1	TPIS_ENTHI	O02611 entamoeba h
45	28	68.3	284	1	YN60_YEAST	P42840 saccharomyc

ALIGNMENTS

RESULT 1
CDN1_FELCA
ID CDN1_FELCA STANDARD; PRT; 164 AA.
AC O19002;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cyclin-dependent kinase inhibitor 1 (p21) (CDK-interacting protein 1).
GN CDKN1A OR CIP1 OR WAF1.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph node;
RX MEDLINE=98036042; PubMed=9370275;
RA Okuda M., Minehata K., Setoguchi A., Cho K.-W., Nakamura N.,
RA Nishigaki K., Watari T., Cevario S., O'Brien S.J., Tsujimoto H.,
RA Hasegawa A.;
RT "Cloning and chromosome mapping of the feline genes p21WAF1 and
RT p27Kip1.";
RL Gene 198:141-147(1997).
CC -!- FUNCTION: MAY BE THE IMPORTANT INTERMEDIATE BY WHICH P53 MEDIATES
CC ITS ROLE AS AN INHIBITOR OF CELLULAR PROLIFERATION IN RESPONSE TO
CC DNA DAMAGE. MAY BIND TO AND INHIBIT CYCLIN-DEPENDENT KINASE
CC ACTIVITY, PREVENTING PHOSPHORYLATION OF CRITICAL CYCLIN-DEPENDENT
CC KINASE SUBSTRATES AND BLOCKING CELL CYCLE PROGRESSION (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: THE N-TERMINAL OF CIP1 AND KIP ARE SIMILAR.
CC --

CC --
DR EMBL; D84650; BAA23168.1; -.
DR InterPro; IPR003175; CDI.
DR Pfam; PF02234; CDI; 1.
KW Cell cycle; Nuclear protein; Zinc-finger.
FT ZN_FING 13 41 C4-TYPE (POTENTIAL).
FT DOMAIN 141 156 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 164 AA; 18315 MW; 0F7912A76C78BF38 CRC64;

Query Match 92.7%; Score 38; DB 1; Length 164;
Best Local Similarity 87.5%; Pred. No. 0.11;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HAKRRLIF 8
|:||||||
Db 152 HSKRRLIF 159

```
RESULT 2
CDN1_HUMAN
ID   CDN1_HUMAN     STANDARD;      PRT;   164 AA.
AC   P38936; Q9BUT4;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cyclin-dependent kinase inhibitor 1 (p21) (CDK-interacting protein 1)
DE   (Melanoma differentiation associated protein 6) (MDA-6).
GN   CDKN1A OR CDKN1 OR CIP1 OR WAF1 OR MDA6 OR SDI1 OR PIC1 OR CAP20.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94061996; PubMed=8242751;
RA   Harper J.W., Adami G.R., Wei N., Keyomarsi K., Elledge S.J.;
RT   "The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1
RT   cyclin-dependent kinases.";
RL   Cell 75:805-816(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94061997; PubMed=8242752;
RA   El-Deiry W.S., Tokino T., Velculescu V.E., Levy D.B., Parsons R.,
RA   Trent J.M., Lin D., Mercer W.E., Kinzler K.W., Vogelstein B.;
RT   "WAF1, a potential mediator of p53 tumor suppression.";
RL   Cell 75:817-825(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94081955; PubMed=8259214;
RA   Xiong Y., Hannon G.J., Zhang H., Casso D., Kobayashi R., Beach D.;
RT   "p21 is a universal inhibitor of cyclin kinases.";
RL   Nature 366:701-704(1993).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Jiang H., Fisher P.B.;
RT   "Use of a sensitive and efficient subtraction hybridization protocol
RT   for the identification of genes differentially regulated during the
RT   induction of differentiation in human melanoma cells.";
RL   Mol. Cell. Differ. 1:285-299(1993).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Jiang H., Lin J., Herlyn M., Kerbel R.S., Weissman B.E.,
RA   Welch D.R., Fisher P.B.;
RL   Submitted (MAY-1994) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94170884; PubMed=8125163;
RA   Noda A., Ning Y., Venable S.F., Pereira-Smith O.M., Smith J.R.;
RT   "Cloning of senescent cell-derived inhibitors of DNA synthesis using
RT   an expression screen.";
RL   Exp. Cell Res. 211:90-98(1994).
RN   [7]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95384154; PubMed=7655464;
RA   Mousses S., Oezcelik H., Lee P.D., Malkin D., Bull S.B.,
RA   Andrulis I.L.;
RT   "Two variants of the CIP1/WAF1 gene occur together and are associated
RT   with human cancer.";
RL   Hum. Mol. Genet. 4:1089-1092(1995).
RN   [8]
RP   SEQUENCE FROM N.A., AND VARIANT ARG-31.
RA   Rieder M.J., Braun A.C., Montoya M.A., Chung M.-W., Nguyen C.P.,
RA   Nguyen D.A., Livingston R.J., Poel C.L., Robertson P.D.,
RA   Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RA   Palmer S.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A., AND VARIANT ARG-31.
RC   TISSUE=Eye, and Lung;
RA   Strausberg R.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 139-160.
RX   MEDLINE=97015085; PubMed=8861913;
RA   Gulbis J.M., Kelman Z., Hurwitz J., O'Donnell M., Kuriyan J.;
RT   "Structure of the C-terminal region of p21(WAF1/CIP1) complexed with
RT   human PCNA.";
RL   Cell 87:297-306(1996).
CC   -!- FUNCTION: MAY BE THE IMPORTANT INTERMEDIATE BY WHICH P53 MEDIATES
CC       ITS ROLE AS AN INHIBITOR OF CELLULAR PROLIFERATION IN RESPONSE TO
CC       DNA DAMAGE. MAY BIND TO AND INHIBIT CYCLIN-DEPENDENT KINASE
CC       ACTIVITY, PREVENTING PHOSPHORYLATION OF CRITICAL CYCLIN-DEPENDENT
CC       KINASE SUBSTRATES AND BLOCKING CELL CYCLE PROGRESSION.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: IS EXPRESSED IN ALL ADULT HUMAN TISSUES,
CC       WITH 5-FOLD LOWER LEVELS OBSERVED IN THE BRAIN.
CC   -!- INDUCTION: BY P53, MEZEREIN (ANTILEUKEMIC COMPOUND) AND INTERFERON
CC       BETA.
CC   -!- SIMILARITY: THE N-TERMINAL OF CIP1 AND KIP ARE SIMILAR.
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/CDKN1AID139.html".
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L25610; AAA16109.1; -.
DR   EMBL; S67388; AAB29246.1; -.
DR   EMBL; U09579; AAA85641.1; -.
DR   EMBL; U03106; AAC04313.1; -.
DR   EMBL; L26165; AAA19811.1; -.
DR   EMBL; L47233; AAB59560.1; ALT_INIT.
DR   EMBL; AF497972; AAM11787.1; -.
DR   EMBL; Z85996; CAB06656.1; -.
DR   EMBL; BC000275; AAH00275.1; -.
DR   EMBL; BC000312; AAH00312.1; -.
DR   EMBL; BC001935; AAH01935.1; -.
DR   EMBL; BC013967; AAH13967.1; -.
DR   PIR; S39357; S39357.
DR   SWISS-2DPAGE; P38936; HUMAN.
DR   Genew; HGNC:1784; CDKN1A.
DR   MIM; 116899; -.
DR   InterPro; IPR003175; CDI.
DR   Pfam; PF02234; CDI; 1.
KW   Cell cycle; Nuclear protein; Zinc-finger; Polymorphism.
FT   ZN_FING      13     41       C4-TYPE (POTENTIAL).
FT   DOMAIN      141    156       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   VARIANT      31     31       S -> R (IN DBSNP:1801270).
FT                                /FTId=VAR_011870.
SQ   SEQUENCE   164 AA;  18119 MW;  98D1E7C519ADFCA9 CRC64;

  Query Match             92.7%;  Score 38;  DB 1;  Length 164;
  Best Local Similarity   87.5%;  Pred. No. 0.11;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 HAKRRLIF 8
            |:|||||
Db     152 HSKRRLIF 159

RESULT 3
CDN1_MOUSE
ID   CDN1_MOUSE     STANDARD;      PRT;   159 AA.
AC   P39689;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
```

```
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cyclin-dependent kinase inhibitor 1 (p21) (CDK-interacting protein 1)
DE   (Melanoma differentiation associated protein).
GN   CDKN1A OR CIP1 OR WAF1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BXSB; TISSUE=Spleen;
RX   MEDLINE=94366751; PubMed=8084607;
RA   Huppi K., Siwarski D., Dosik J., Michieli P., Chedid M., Reed S.,
RA   Mock B., Givol D., Mushinski J.F.;
RT   "Molecular cloning, sequencing, chromosomal localization and
RT   expression of mouse p21 (Waf1).";
RL   Oncogene 9:3017-3020(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95316868; PubMed=7796420;
RA   El-Deiry W.S., Tokino T., Waldman T., Velculescu V., Oliner J.D.,
RA   Burell M., Hill D.E., Rees J.L., Hamilton S.R., Kinzler K.W.,
RA   Vogelstein B.;
RT   "Topological control of p21WAF1/CIP1 expression in normal and
RT   neoplastic tissues.";
RL   Cancer Res. 55:2910-2919(1995).
RN   [3]
RP   SEQUENCE OF 1-143 FROM N.A.
RX   MEDLINE=94061997; PubMed=8242752;
RA   El-Deiry W.S., Tokino T., Velculescu V.E., Levy D.B., Parsons R.,
RA   Trent J.M., Lin D., Mercer W.E., Kinzler K.W., Vogelstein B.;
RT   "WAF1, a potential mediator of p53 tumor suppression.";
RL   Cell 75:817-825(1993).
CC   -!- FUNCTION: MAY BE THE IMPORTANT INTERMEDIATE BY WHICH P53 MEDIATES
CC       ITS ROLE AS AN INHIBITOR OF CELLULAR PROLIFERATION IN RESPONSE TO
CC       DNA DAMAGE. MAY BIND TO AND INHIBIT CYCLIN-DEPENDENT KINASE
CC       ACTIVITY, PREVENTING PHOSPHORYLATION OF CRITICAL CYCLIN-DEPENDENT
CC       KINASE SUBSTRATES AND BLOCKING CELL CYCLE PROGRESSION.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- INDUCTION: BY P53, MEZEREIN (ANTILEUKEMIC COMPOUND) AND INTERFERON
CC       BETA.
CC   -!- SIMILARITY: THE N-TERMINAL OF CIP1 AND KIP ARE SIMILAR.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U09507; AAB60456.1; -.
DR   EMBL; U24173; AAC52220.1; -.
DR   PIR; A49438; A49438.
DR   MGD; MGI:104556; Cdkn1a.
DR   InterPro; IPR003175; CDI.
DR   Pfam; PF02234; CDI; 1.
KW   Cell cycle; Nuclear protein; Zinc-finger.
FT   ZN_FING      12     40       C4-TYPE (POTENTIAL).
FT   CONFLICT     30     30       R -> S (IN REF. 3).
FT   CONFLICT     56     57       TP -> RQ (IN REF. 3).
SQ   SEQUENCE   159 AA;  17785 MW;  37B7C22B9A2FD089 CRC64;

  Query Match             90.2%;  Score 37;  DB 1;  Length 159;
  Best Local Similarity   75.0%;  Pred. No. 0.18;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          |:||||:|
Db    147 HSKRRLVF 154


RESULT 4
```

```
RPN4_YEAST
ID   RPN4_YEAST     STANDARD;      PRT;   531 AA.
AC   Q03465;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   26S proteasome regulatory subunit RPN4 (Nuclear protein SON1) (UB
DE   fusion degradation protein 5).
GN   RPN4 OR SON1 OR UFD5 OR YDL020C OR D2840.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W303;
RX   MEDLINE=93292918; PubMed=8514125;
RA   Nelson M.K., Kurihara T., Silver P.A.;
RT   "Extragenic suppressors of mutations in the cytoplasmic C terminus of
RT   SEC63 define five genes in Saccharomyces cerevisiae.";
RL   Genetics 134:159-173(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RA   Andre B., Vissers S., Urrestarazu L.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   CHARACTERIZATION.
RC   STRAIN=S288c;
RX   MEDLINE=95340540; PubMed=7615550;
RA   Johnson E.S., Ma P.C.M., Ota I.M., Varshavsky A.;
RT   "A proteolytic pathway that recognizes ubiquitin as a degradation
RT   signal.";
RL   J. Biol. Chem. 270:17442-17456(1995).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=98171302; PubMed=9512348;
RA   Fujimuro M., Tanaka K., Yokosawa H., Toh-E A.;
RT   "Son1p is a component of the 26S proteasome of the yeast
RT   Saccharomyces cerevisiae.";
RL   FEBS Lett. 423:149-154(1998).
CC   -!- FUNCTION: MAY PLAY A ROLE IN NUCLEAR INTEGRITY, IS REQUIRED FOR
CC       NORMAL GROWTH AT LOW TEMPERATURES. SON1 MUTANTS GROW SLOWLY AT LOW
CC       TEMPERATURES AND SHOW PARTIAL MISLOCALIZATION OF NUCLEAR ANTIGENS.
CC       PROBABLY INTERACTS WITH SEC63. COMPONENT OF 26S PROTEASOME.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L00928; AAA35067.1; -.
DR   EMBL; Z48432; CAA88339.1; -.
DR   EMBL; Z74068; CAA98579.1; -.
DR   PIR; S30806; S30806.
DR   PIR; S41986; S41986.
DR   TRANSFAC; T04539; -.
DR   SGD; S0002178; RPN4.
DR   InterPro; IPR000822; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 2.
DR   SMART; SM00355; ZnF_C2H2; 1.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 1.
KW   Proteasome; Nuclear protein.
FT   DOMAIN      211    229       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      300    315       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      382    398       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ   SEQUENCE   531 AA;  60152 MW;  4316281AC09FBE7F CRC64;

  Query Match             85.4%;  Score 35;  DB 1;  Length 531;
```

```
  Best Local Similarity   62.5%;  Pred. No. 1.9;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          ||||:::|
Db    468 HAKRKIVF 475


RESULT 5
MT04_STAAU
ID   MT04_STAAU     STANDARD;      PRT;   214 AA.
AC   Q9KWZ4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical methyltransferase (EC 2.1.1.-).
OS   Staphylococcus aureus.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COL;
RX   MEDLINE=20031141; PubMed=10566865;
RA   de Lencastre H., Wu S.W., Pinho M.G., Ludovice A.M., Filipe S.,
RA   Gardete S., Sobral R., Gill S., Chung M., Tomasz A.;
RT   "Antibiotic resistance as a stress response: complete sequencing of a
RT   large number of chromosomal loci in Staphylococcus aureus strain COL
RT   that impact on the expression of resistance to methicillin.";
RL   Microb. Drug Resist. 5:163-175(1999).
CC   -!- FUNCTION: PROBABLE METHYLTRANSFERASE.
CC   -!- SIMILARITY: BELONGS TO THE UPF0155 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Y14816; CAB82477.1; -.
DR   InterPro; IPR004395; Cons_hypoth91.
DR   InterPro; IPR003358; Methyltransf_4.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF02390; Methyltransf_4; 1.
DR   TIGRFAMs; TIGR00091; Cons_hypoth91; 1.
KW   Hypothetical protein; Transferase; Methyltransferase.
SQ   SEQUENCE   214 AA;  25275 MW;  789C314D41B2CC68 CRC64;

  Query Match             80.5%;  Score 33;  DB 1;  Length 214;
  Best Local Similarity   75.0%;  Pred. No. 2.1;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          ||||||  :
Db    124 HAKRRLTY 131


RESULT 6
MT04_MYCHO
ID   MT04_MYCHO     STANDARD;      PRT;   202 AA.
AC   Q9F411;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical methyltransferase MG347 homolog (EC 2.1.1.-).
OS   Mycoplasma hominis.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2098;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PG21;
RX   MEDLINE=20448743; PubMed=10991851;
RA   Bebear C.M., Grau O., Charron A., Renaudin H., Gruson D., Bebear C.;
RT   "Cloning and nucleotide sequence of the DNA gyrase (gyrA) gene from
RT   Mycoplasma hominis and characterization of quinolone-resistant mutants
RT   selected In vitro with trovafloxacin.";
RL   Antimicrob. Agents Chemother. 44:2719-2727(2000).
CC   -!- FUNCTION: PROBABLE METHYLTRANSFERASE.
CC   -!- SIMILARITY: BELONGS TO THE UPF0155 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U59880; AAG28841.1; -.
DR   InterPro; IPR004395; Cons_hypoth91.
DR   InterPro; IPR003358; Methyltransf_4.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF02390; Methyltransf_4; 1.
DR   TIGRFAMs; TIGR00091; Cons_hypoth91; 1.
KW   Hypothetical protein; Transferase; Methyltransferase.
SQ   SEQUENCE   202 AA;  23951 MW;  D8993C44B381BB0B CRC64;

  Query Match             75.6%;  Score 31;  DB 1;  Length 202;
  Best Local Similarity   62.5%;  Pred. No. 5.8;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          | ||||::
Db    114 HYKRRLVY 121


RESULT 7
HETR_ANASP
ID   HETR_ANASP     STANDARD;      PRT;   299 AA.
AC   P27709;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Heterocyst differentiation control protein.
GN   HETR OR ALR2339.
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91138965; PubMed=1840555;
RA   Buikema W.J., Haselkorn R.;
RT   "Characterization of a gene controlling heterocyst differentiation in
RT   the cyanobacterium Anabaena 7120.";
RL   Genes Dev. 5:321-330(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21595285; PubMed=11759840;
RA   Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA   Yasuda M., Tabata S.;
RT   "Complete genomic sequence of the filamentous nitrogen-fixing
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   DNA Res. 8:205-213(2001).
CC   -!- FUNCTION: CONTROLS HETEROCYST DIFFERENTIATION.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ONLY IN THE CELLS THAT ARE GOING
CC       TO DIFFERENTIATE INTO HETEROCYSTS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M37779; AAA21998.1; -.
DR   EMBL; AP003589; BAB74038.1; -.
DR   PIR; A38705; A38705.
DR   MEROPS; S48.001; -.
DR   InterPro; IPR005319; Peptidase_S48.
DR   Pfam; PF03574; Peptidase_S48; 1.
KW   Heterocyst; Complete proteome.
FT   VARIANT     179    179       S -> N (IN STRAIN 216, UNABLE TO CONTROL
FT                                HETEROCYST DIFFERENTIATION).
SQ   SEQUENCE   299 AA;  34969 MW;  74AF6CA9D87F9126 CRC64;

  Query Match             75.6%;  Score 31;  DB 1;  Length 299;
  Best Local Similarity   62.5%;  Pred. No. 8.9;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 HAKRRLIF 8
           | ||||:: 
Db     185 HIKRRLLY 192


RESULT 8
PHOQ_SALTY
ID   PHOQ_SALTY     STANDARD;      PRT;   487 AA.
AC   P14147; Q9L3L1;
DT   01-JAN-1990 (Rel. 13, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Virulence sensor protein phoQ (EC 2.7.3.-).
GN   PHOQ OR STM1230.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 10428;
RX   MEDLINE=89296942; PubMed=2544889;
RA   Miller S.I., Kukral A.M., Mekalanos J.J.;
RT   "A two-component regulatory system (phoP phoQ) controls Salmonella
RT   typhimurium virulence.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:5054-5058(1989).
RN   [2]
RP   SEQUENCE FROM N.A., AND MUTAGENESIS OF ARG-313.
RC   STRAIN=SL1344;
RX   MEDLINE=21437654; PubMed=11553591;
RA   Cano D.A., Martinez-Moya M., Pucciarelli M.G., Groisman E.A.,
RA   Casadesus J., Garcia-del Portillo F.;
RT   "Salmonella enterica serovar Typhimurium response involved in
RT   attenuation of pathogen intracellular proliferation.";
RL   Infect. Immun. 69:6463-6474(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- FUNCTION: Member of the two-component regulatory system phoQ/phoP
CC       which regulates the expression of genes involved in virulence and
CC       promotes intramacrophage survival of S.typhimurium. Is required to
CC       attenuate bacterial growth within fibroblast cells. PhoQ may
CC       function as a membrane-associated protein kinase that
CC       phosphorylates phoP in response to environmental signals.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Probable).
CC   -!- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M24424; AAA27189.1; -.
DR   EMBL; AJ272210; CAB75592.1; -.
DR   EMBL; AE008753; AAL20159.1; -.
DR   PIR; B32932; VZEBPT.
DR   StyGene; SG10294; phoQ.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR003660; HAMP.
DR   InterPro; IPR004359; HIS_KIN_sig.
DR   Pfam; PF00512; signal; 2.
DR   Pfam; PF00672; HAMP; 1.
DR   Pfam; PF02518; HATPase_c; 2.
DR   SMART; SM00387; HATPase_c; 1.
DR   PROSITE; PS50109; HIS_KIN; 1.
KW   Sensory transduction; Transferase; Kinase; Phosphorylation;
KW   Transmembrane; Inner membrane; Growth regulation; Virulence;
KW   Complete proteome.
FT   DOMAIN        1     16       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     17     37       POTENTIAL.
FT   DOMAIN       38    194       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    195    215       POTENTIAL.
FT   DOMAIN      216    487       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      256    451       HISTIDINE KINASE.
FT   MOD_RES     277    277       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   VARIANT      82     99       MISSING (IN STRAIN ATCC 10428).
FT   VARIANT     442    459       MISSING (IN STRAIN ATCC 10428).
FT   MUTAGEN     313    313       R->W: INCREASED ABILITY TO PROLIFERATE
FT                                WITHIN FIBROBLASTS.
SQ   SEQUENCE   487 AA;  55466 MW;  BDCFEFC56F4CA058 CRC64;

  Query Match             75.6%;  Score 31;  DB 1;  Length 487;
  Best Local Similarity   62.5%;  Pred. No. 15;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 HAKRRLIF 8
             |:|| |:| 
Db     423 HSKRSLVF 430


RESULT 9
YEF5_YEAST
ID   YEF5_YEAST     STANDARD;      PRT;   141 AA.
AC   P32616;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   Hypothetical 16.5 kDa protein in GLY1-GDA1 intergenic region.
GN   YEL045C OR SYGP-ORF33.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mulligan J.T., Dietrich F.S., Hennessey K.M., Sehl P., Komp C.,
RA   Wei Y., Taylor P., Nakahara K., Roberts D., Davis R.W.;
RL   Submitted (FEB-1993) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA   Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA   Chung E., Duncan M., Guzman E., Hartzell G., Hunicke-Smith S.,
RA   Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
```

```
RA   Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA   Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA   Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U18779; AAB64997.1; -.
DR   PIR; S30832; S30832.
DR   SGD; S0000771; YEL045C.
KW   Hypothetical protein; ATP-binding; Transmembrane.
FT   NP_BIND      15     22       ATP (POTENTIAL).
FT   TRANSMEM     38     58       POTENTIAL.
FT   TRANSMEM     67     87       POTENTIAL.
SQ   SEQUENCE   141 AA;  16468 MW;  F6604AC5343A5D5C CRC64;

  Query Match             73.2%;  Score 30;  DB 1;  Length 141;
  Best Local Similarity   75.0%;  Pred. No. 6.7;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          |||| | |
Db      4 HAKRTLAF 11


RESULT 10
MT04_TREPA
ID   MT04_TREPA     STANDARD;      PRT;   250 AA.
AC   O83477;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical methyltransferase TP0464 (EC 2.1.1.-).
GN   TP0464.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Nichols;
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
CC   -!- FUNCTION: PROBABLE METHYLTRANSFERASE.
CC   -!- SIMILARITY: BELONGS TO THE UPF0155 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001223; AAC65448.1; -.
DR   TIGR; TP0464; -.
DR   InterPro; IPR004395; Cons_hypoth91.
DR   InterPro; IPR003358; Methyltransf_4.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF02390; Methyltransf_4; 1.
DR   TIGRFAMs; TIGR00091; Cons_hypoth91; 1.
KW   Hypothetical protein; Transferase; Methyltransferase;
KW   Complete proteome.
SQ   SEQUENCE   250 AA;  28068 MW;  C5A227F4E4D15CC0 CRC64;

  Query Match             73.2%;  Score 30;  DB 1;  Length 250;
  Best Local Similarity   62.5%;  Pred. No. 12;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          | ||||::
Db    168 HHKRRLLY 175


RESULT 11
YK09_YEAST
ID   YK09_YEAST     STANDARD;      PRT;   751 AA.
AC   P36124;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical 85.5 kDa protein in SAP190-SPO14 intergenic region.
GN   YKR029C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Urrestarazu L.A., Jauniaux J.-C.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO YEAST YJL105W AND S.POMBE SPAC22E12.11C.
CC   -!- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC   -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z28254; CAA82101.1; -.
DR   PIR; S38101; S38101.
DR   SGD; S0001737; YKR029C.
DR   InterPro; IPR001214; SET.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF00628; PHD; 1.
DR   Pfam; PF00856; SET; 1.
DR   SMART; SM00249; PHD; 1.
DR   SMART; SM00317; SET; 1.
DR   PROSITE; PS50280; SET; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
KW   Hypothetical protein; Zinc-finger.
FT   ZN_FING      117    166       PHD-TYPE.
FT   DOMAIN       334    460       SET.
SQ   SEQUENCE   751 AA;  85479 MW;  934621768C36230B CRC64;

  Query Match             73.2%;  Score 30;  DB 1;  Length 751;
  Best Local Similarity   85.7%;  Pred. No. 41;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 AKRRLIF 8
          ||||:||
Db    376 AKRRVIF 382


RESULT 12
H4_ENTHI
ID   H4_ENTHI       STANDARD;      PRT;   118 AA.
```

```
AC   P40287;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Histone H4.
OS   Entamoeba histolytica.
OC   Eukaryota; Entamoebidae; Entamoeba.
OX   NCBI_TaxID=5759;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HM-1:IMSS;
RA   Tanaka T.;
RL   Submitted (XXX-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SFL-3;
RX   MEDLINE=96065756; PubMed=7477106;
RA   Binder M., Ortner S., Plaimauer B., Fodinger M., Wiedermann G.,
RA   Scheiner O., Duchene M.;
RT   "Sequence and organization of an unusual histone H4 gene in the human
RT   parasite Entamoeba histolytica.";
RL   Mol. Biochem. Parasitol. 71:243-247(1995).
CC   -!- FUNCTION: HISTONE H4, ALONG WITH HISTONE H3, PLAYS A CENTRAL ROLE
CC       IN NUCLEOSOME FORMATION.
CC   -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC       H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H4 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L35898; AAB67323.1; -.
DR   EMBL; X84010; CAA58833.1; -.
DR   EMBL; X84009; CAA58831.1; -.
DR   InterPro; IPR001951; Histone_H4.
DR   InterPro; IPR004822; Histone_core.
DR   Pfam; PF00125; histone; 1.
DR   PRINTS; PR00623; HISTONEH4.
DR   ProDom; PD001827; Histone_H4; 1.
DR   SMART; SM00417; H4; 1.
DR   PROSITE; PS00047; HISTONE_H4; FALSE_NEG.
KW   Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding.
SQ   SEQUENCE   118 AA;  12870 MW;  D804259C36A81603 CRC64;

  Query Match             70.7%;  Score 29;  DB 1;  Length 118;
  Best Local Similarity   71.4%;  Pred. No. 9.5;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLI 7
          ||||| :
Db     92 HAKRRTV 98


RESULT 13
FTN_HELPJ
ID   FTN_HELPJ      STANDARD;      PRT;   167 AA.
AC   Q9ZLI1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nonheme iron-containing ferritin.
GN   PFR OR JHP0598.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- FUNCTION: IRON-STORAGE PROTEIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE FERRITIN FAMILY. PROKARYOTIC SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001491; AAD06160.1; -.
DR   HSSP; P23887; 1EUM.
DR   InterPro; IPR001519; Ferritin.
DR   Pfam; PF00210; ferritin; 1.
KW   Iron storage; Complete proteome.
FT   METAL        17     17       IRON (BY SIMILARITY).
SQ   SEQUENCE   167 AA;  19314 MW;  D18B7F3F2CAD9CFC CRC64;

  Query Match             70.7%;  Score 29;  DB 1;  Length 167;
  Best Local Similarity   62.5%;  Pred. No. 14;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          |||: :||
Db     53 HAKKLIIF 60


RESULT 14
TPIS_AEDTO
ID   TPIS_AEDTO     STANDARD;      PRT;   206 AA.
AC   P92119;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Triosephosphate isomerase (EC 5.3.1.1) (TIM) (Fragment).
GN   TPI.
OS   Aedes togoi (Mosquito).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC   Culicoidea; Ochlerotatus.
OX   NCBI_TaxID=55967;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97438232; PubMed=9294007;
RA   Tyshenko M.G., Walker V.K.;
RT   "Towards a reconciliation of the introns early or late views:
RT   triosephosphate isomerase genes from insects.";
RL   Biochim. Biophys. Acta 1353:131-136(1997).
CC   -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glycerone
CC       phosphate.
CC   -!- PATHWAY: PLAYS AN IMPORTANT ROLE IN SEVERAL METABOLIC PATHWAYS.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; U82708; AAB48449.1; -.
DR   HSSP; P00940; 1TPH.
DR   InterPro; IPR000652; Triophos_ismrse.
DR   Pfam; PF00121; TIM; 1.
DR   ProDom; PD001005; Triophos_ismrse; 1.
DR   TIGRFAMs; TIGR00419; tim; 1.
DR   PROSITE; PS00171; TIM; 1.
KW   Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
KW   Pentose shunt.
FT   NON_TER       1      1
FT   ACT_SITE     76     76       BY SIMILARITY.
FT   ACT_SITE    146    146       BY SIMILARITY.
FT   NON_TER     206    206
SQ   SEQUENCE   206 AA;  21925 MW;  AD89BBA000FF8B69 CRC64;

  Query Match             70.7%;  Score 29;  DB 1;  Length 206;
  Best Local Similarity   62.5%;  Pred. No. 17;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          |::|| ||
Db     76 HSERRAIF 83


RESULT 15
TPIS_ANOME
ID   TPIS_ANOME     STANDARD;      PRT;   206 AA.
AC   P91895;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Triosephosphate isomerase (EC 5.3.1.1) (TIM) (Fragment).
GN   TPI.
OS   Anopheles merus (Mosquito).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC   Culicoidea; Anopheles.
OX   NCBI_TaxID=30066;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97438232; PubMed=9294007;
RA   Tyshenko M.G., Walker V.K.;
RT   "Towards a reconciliation of the introns early or late views:
RT   triosephosphate isomerase genes from insects.";
RL   Biochim. Biophys. Acta 1353:131-136(1997).
CC   -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glycerone
CC       phosphate.
CC   -!- PATHWAY: PLAYS AN IMPORTANT ROLE IN SEVERAL METABOLIC PATHWAYS.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U82707; AAB48448.1; -.
DR   HSSP; P00940; 1TPH.
DR   InterPro; IPR000652; Triophos_ismrse.
DR   Pfam; PF00121; TIM; 1.
DR   ProDom; PD001005; Triophos_ismrse; 1.
DR   TIGRFAMs; TIGR00419; tim; 1.
DR   PROSITE; PS00171; TIM; 1.
KW   Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
KW   Pentose shunt.
FT   NON_TER       1      1
FT   ACT_SITE     76     76       BY SIMILARITY.
FT   ACT_SITE    146    146       BY SIMILARITY.
FT   NON_TER     206    206
SQ   SEQUENCE   206 AA;  21878 MW;  049162C164094385 CRC64;

  Query Match             70.7%;  Score 29;  DB 1;  Length 206;
  Best Local Similarity   62.5%;  Pred. No. 17;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          |::|| ||
Db     76 HSERRAIF 83


Search completed: December 14, 2002, 15:46:44
Job time : 24.5 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: December 14, 2002, 13:14:29 ; Search time 57 Seconds
(without alignments)
18.702 Million cell updates/sec

Title: US-09-726-470A-35
Perfect score: 41
Sequence: 1 HAKRRLIF 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	100.0	8	22	AAG65137	Synthetic peptide,
2	41	100.0	8	22	AAG65150	p21 derived cyclin
3	41	100.0	11	22	AAG65147	p21 derived cyclin
4	41	100.0	12	22	AAG65100	p21WAFI C-terminal
5	41	100.0	20	17	AAR99664	p21WAF1 peptide 10
6	40	97.6	8	22	AAU05708	p21 C-terminus der
7	39	95.1	8	22	AAU05702	p21 C-terminus der
8	39	95.1	8	22	AAU05707	p21 C-terminus der
9	39	95.1	8	22	AAU05742	p21 C-terminus der
10	38	92.7	8	22	AAG65127	p21WAFI C-terminal
11	38	92.7	8	22	AAG66266	p21 C-terminus der
12	38	92.7	8	22	AAU05703	p21 C-terminus der
13	38	92.7	8	22	AAU05714	p21 C-terminus der
14	38	92.7	9	22	AAG65110	p21WAFI C-terminal
15	38	92.7	9	22	AAG65122	p21WAFI C-terminal
16	38	92.7	10	22	AAG65109	p21WAFI C-terminal
17	38	92.7	10	22	AAG65117	p21WAFI C-terminal
18	38	92.7	11	22	AAG65108	p21WAFI C-terminal
19	38	92.7	11	22	AAG65111	p21WAFI C-terminal
20	38	92.7	12	22	AAG65090	p21WAFI C-terminal
21	38	92.7	12	22	AAG65091	p21WAFI C-terminal
22	38	92.7	12	22	AAG65092	p21WAFI C-terminal
23	38	92.7	12	22	AAG65093	p21WAFI C-terminal
24	38	92.7	12	22	AAG65094	p21WAFI C-terminal
25	38	92.7	12	22	AAG65096	p21WAFI C-terminal
26	38	92.7	12	22	AAG65097	p21WAFI C-terminal
27	38	92.7	12	22	AAG65098	p21WAFI C-terminal
28	38	92.7	12	22	AAG65107	p21WAFI C-terminal
29	38	92.7	16	18	AAW44237	Human p21waf1 frag
30	38	92.7	20	17	AAR98310	p21WAF1 peptide 10
31	38	92.7	20	17	AAR98311	p21WAF1 peptide 11
32	38	92.7	20	17	AAR99652	p21WAF1 peptide 10
33	38	92.7	20	17	AAR99654	p21WAF1 peptide 10
34	38	92.7	20	17	AAR99655	p21WAF1 peptide 10
35	38	92.7	20	17	AAR99657	p21WAF1 peptide 10
36	38	92.7	20	17	AAR99659	p21WAF1 peptide 10
37	38	92.7	20	17	AAR99660	p21WAF1 peptide 10
38	38	92.7	20	17	AAR99661	p21WAF1 peptide 10
39	38	92.7	20	17	AAR99662	p21WAF1 peptide 10
40	38	92.7	20	17	AAR99671	p21WAF1 peptide 10
41	38	92.7	20	17	AAR99658	p21WAF1 peptide 10
42	38	92.7	20	18	AAW44221	Human p21 fragment
43	38	92.7	20	18	AAW44222	Human p21 fragment
44	38	92.7	20	18	AAW44223	Human p21 fragment
45	38	92.7	20	21	AAB17272	p16-mimetic peptid

ALIGNMENTS

RESULT 1
AAG65137
ID AAG65137 standard; Peptide; 8 AA.
XX
AC AAG65137;
XX
DT 21-NOV-2001 (first entry)
XX
DE Synthetic peptide, p21 C-terminus (S153A).
XX
KW Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW inhibitor; proliferative disorder; cancer; leukaemia;
KW drug screening; p21 C-terminus (S153A).
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "Optional Hydrogenated N-terminus"
FT Modified-site 8
FT /note= "Optional C-terminal carboxamide or amide"
XX
PN WO200140142-A2.
XX
PD 07-JUN-2001.
XX
PF 29-NOV-2000; 2000WO-GB04550.
XX
PR 30-NOV-1999; 99GB-0028323.
XX
PA (CYCL-) CYCLACEL LTD.
XX

```
PI   Zheleva DI,  Fischer PM,  McInnes C,  Andrews MJI,  Chan WC;
PI   Atkinson GE;
XX
DR   WPI; 2001-488493/53.
XX
PT   New p21 derived peptides and their variants, particularly useful as
PT   selective inhibitors of CDK2/cyclin interaction for treating
PT   proliferative disorders e.g. cancers and leukaemias, and in assays for
PT   identifying CDK/cyclin inhibitors -
XX
PS   Claim 15; Page 84; 102pp; English.
XX
CC   The invention relates to peptide and their variants derived from p21WAFI,
CC   which are inhibitors of CDK2 activity by binding to G1 and
CC   S phase specific cyclins which activate CDK2; selective inhibitors of
CC   CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
CC   The variants of the peptide may have further amino acids at either end
CC   or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC   The peptides are specific regions of p21WAFI that bind to G1 and S
CC   phase specific cyclins, preferably cyclins which activate CDK2. One
CC   of the peptides corresponds to p21(149-159). The peptides are used for
CC   treating proliferative disorders, e.g. cancers and leukaemias. The
CC   peptides are also for identifying substances which interfere with
CC   protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC   especially CDK/cyclin interactions, and which are capable of inhibiting
CC   CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC   competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC   may be used to identify substances that bind to, or inhibit peptide-
CC   cyclin interactions. Substances for screening in the assays include
CC   antibody products specific for p21 or cyclin binding regions,
CC   combinatorial libraries and single compound collections. The present
CC   sequence is a synthetic peptide derived from the C-terminus of p21.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 41;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          ||||||||
Db      1 HAKRRLIF 8


RESULT 2
AAG65150
ID   AAG65150 standard; Peptide; 8 AA.
XX
AC   AAG65150;
XX
DT   21-NOV-2001  (first entry)
XX
DE   p21 derived cyclin A binding peptide #2.
XX
KW   Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW   inhibitor; proliferative disorder; cancer; leukaemia;
KW   drug screening; mutant; mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "Hydrogenated N-terminus"
FT   Misc-difference 1
FT                   /note= "Optionally a D-form residue"
FT   Misc-difference 2
FT                   /note= "Optionally a D-form residue"
FT   Misc-difference 3
FT                   /note= "Optionally a D-form residue"
FT   Misc-difference 4
FT                   /note= "Optionally a D-form residue"
FT   Misc-difference 5
FT                   /note= "Optionally a D-form residue"
FT   Misc-difference 7
FT                   /note= "Optionally a D-form residue"
FT   Misc-difference 8
FT                   /note= "Optionally a D-form residue"
FT   Modified-site   8
FT                   /note= "C-terminal amide"
XX
PN   WO200140142-A2.
XX
PD   07-JUN-2001.
XX
PF   29-NOV-2000; 2000WO-GB04550.
XX
PR   30-NOV-1999;   99GB-0028323.
XX
PA   (CYCL-) CYCLACEL LTD.
XX
PI   Zheleva DI,  Fischer PM,  McInnes C,  Andrews MJI,  Chan WC;
PI   Atkinson GE;
XX
DR   WPI; 2001-488493/53.
XX
PT   New p21 derived peptides and their variants, particularly useful as
PT   selective inhibitors of CDK2/cyclin interaction for treating
PT   proliferative disorders e.g. cancers and leukaemias, and in assays for
PT   identifying CDK/cyclin inhibitors -
XX
PS   Example 12; Page 53; 102pp; English.
XX
CC   The invention relates to peptide and their variants derived from p21WAFI,
CC   which are inhibitors of CDK2 activity by binding to G1 and
CC   S phase specific cyclins which activate CDK2; selective inhibitors of
CC   CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
CC   The variants of the peptide may have further amino acids at either end
CC   or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC   The peptides are specific regions of p21WAFI that bind to G1 and S
CC   phase specific cyclins, preferably cyclins which activate CDK2. One
CC   of the peptides corresponds to p21(149-159). The peptides are used for
CC   treating proliferative disorders, e.g. cancers and leukaemias. The
CC   peptides are also for identifying substances which interfere with
CC   protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC   especially CDK/cyclin interactions, and which are capable of inhibiting
CC   CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC   competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC   may be used to identify substances that bind to, or inhibit peptide-
CC   cyclin interactions. Substances for screening in the assays include
CC   antibody products specific for p21 or cyclin binding regions,
CC   combinatorial libraries and single compound collections. The present
CC   sequence is a peptide derived from the C-termius of p21 and used in
CC   a Cyclin A binding experiment, the effect on cyclin A binding of
CC   replacing each residue with its chiral alternative was tested.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 41;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 7.8e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          ||||||||
Db      1 HAKRRLIF 8


RESULT 3
AAG65147
ID   AAG65147 standard; Peptide; 11 AA.
XX
AC   AAG65147;
XX
DT   21-NOV-2001  (first entry)
XX
DE   p21 derived cyclin A binding peptide #1.
```

```
XX
KW   Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW   inhibitor; proliferative disorder; cancer; leukaemia;
KW   drug screening; mutant; mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "Hydrogenated N-terminus"
FT   Misc-difference 5
FT                   /note= "Wild-type Ser substituted by Ala"
FT   Modified-site   11
FT                   /note= "C-terminal amide"
XX
PN   WO200140142-A2.
XX
PD   07-JUN-2001.
XX
PF   29-NOV-2000; 2000WO-GB04550.
XX
PR   30-NOV-1999;   99GB-0028323.
XX
PA   (CYCL-) CYCLACEL LTD.
XX
PI   Zheleva DI,  Fischer PM,  McInnes C,  Andrews MJI,  Chan WC;
PI   Atkinson GE;
XX
DR   WPI; 2001-488493/53.
XX
PT   New p21 derived peptides and their variants, particularly useful as
PT   selective inhibitors of CDK2/cyclin interaction for treating
PT   proliferative disorders e.g. cancers and leukaemias, and in assays for
PT   identifying CDK/cyclin inhibitors -
XX
PS   Example 10; Page 51; 102pp; English.
XX
CC   The invention relates to peptide and their variants derived from p21WAFI,
CC   which are inhibitors of CDK2 activity by binding to G1 and
CC   S phase specific cyclins which activate CDK2; selective inhibitors of
CC   CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
CC   The variants of the peptide may have further amino acids at either end
CC   or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC   The peptides are specific regions of p21WAFI that bind to G1 and S
CC   phase specific cyclins, preferably cyclins which activate CDK2. One
CC   of the peptides corresponds to p21(149-159). The peptides are used for
CC   treating proliferative disorders, e.g. cancers and leukaemias. The
CC   peptides are also for identifying substances which interfere with
CC   protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC   especially CDK/cyclin interactions, and which are capable of inhibiting
CC   CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC   competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC   may be used to identify substances that bind to, or inhibit peptide-
CC   cyclin interactions. Substances for screening in the assays include
CC   antibody products specific for p21 or cyclin binding regions,
CC   combinatorial libraries and single compound collections. The present
CC   sequence is a peptide derived from the C-termius of p21 and used in
CC   a Cyclin A binding experiment.
XX
SQ   Sequence   11 AA;

  Query Match             100.0%;  Score 41;  DB 22;  Length 11;
  Best Local Similarity   100.0%;  Pred. No. 0.069;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 HAKRRLIF 8
           ||||||||
Db       4 HAKRRLIF 11


RESULT 4
AAG65100
ID   AAG65100 standard; Peptide; 12 AA.
XX
AC   AAG65100;
XX
DT   21-NOV-2001  (first entry)
XX
DE   p21WAFI C-terminal synthetic peptide  #5.
XX
KW   Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW   inhibitor; proliferative disorder; cancer; leukaemia;
KW   drug screening; mutant; mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 5
FT                   /note= "Wild-type Ser substituted by Ala"
FT   Modified-site   12
FT                   /note= "Optional C-terminal carboxamide"
XX
PN   WO200140142-A2.
XX
PD   07-JUN-2001.
XX
PF   29-NOV-2000; 2000WO-GB04550.
XX
PR   30-NOV-1999;   99GB-0028323.
XX
PA   (CYCL-) CYCLACEL LTD.
XX
PI   Zheleva DI,  Fischer PM,  McInnes C,  Andrews MJI,  Chan WC;
PI   Atkinson GE;
XX
DR   WPI; 2001-488493/53.
XX
PT   New p21 derived peptides and their variants, particularly useful as
PT   selective inhibitors of CDK2/cyclin interaction for treating
PT   proliferative disorders e.g. cancers and leukaemias, and in assays for
PT   identifying CDK/cyclin inhibitors -
XX
PS   Claim 15; Page 84; 102pp; English.
XX
CC   The invention relates to peptide and their variants derived from p21WAFI,
CC   which are inhibitors of CDK2 activity by binding to G1 and
CC   S phase specific cyclins which activate CDK2; selective inhibitors of
CC   CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
CC   The variants of the peptide may have further amino acids at either end
CC   or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC   The peptides are specific regions of p21WAFI that bind to G1 and S
CC   phase specific cyclins, preferably cyclins which activate CDK2. One
CC   of the peptides corresponds to p21(149-159). The peptides are used for
CC   treating proliferative disorders, e.g. cancers and leukaemias. The
CC   peptides are also for identifying substances which interfere with
CC   protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC   especially CDK/cyclin interactions, and which are capable of inhibiting
CC   CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC   competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC   may be used to identify substances that bind to, or inhibit peptide-
CC   cyclin interactions. Substances for screening in the assays include
CC   antibody products specific for p21 or cyclin binding regions,
CC   combinatorial libraries and single compound collections. The present
CC   sequence is a peptide corresponding to p21(145-164) or a peptide
CC   derived from that region.
XX
SQ   Sequence   12 AA;

  Query Match             100.0%;  Score 41;  DB 22;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 0.075;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 HAKRRLIF 8
           ||||||||
```

```
Db      4 HAKRRLIF 11


RESULT 5
AAR99664
ID   AAR99664 standard; peptide; 20 AA.
XX
AC   AAR99664;
XX
DT   24-MAR-1997  (first entry)
XX
DE   p21WAF1 peptide 10 analogue (#56).
XX
KW   p21WAF1; transcription; tumour suppressor; p53; inhibitor;
KW   cyclin dependent kinase; Cdk;  G1; S phase; cell cycle;
KW   proliferating cell nuclear antigen; PCNA; DNA replication;
KW   processivity factor; polymerase delta; p53-mediated suppression;
KW   proliferation; treatment; hyperproliferative disease; cancer; psoriasis.
XX
OS   Synthetic.
XX
PN   WO9614334-A.
XX
PD   17-MAY-1996.
XX
PF   03-NOV-1995;   95WO-GB02583.
XX
PR   03-NOV-1994;   94GB-0022175.
XX
PA   (UYDU-) UNIV DUNDEE.
XX
PI   Cox LS,  Glover DM,  Lane DP,  Warbrick E;
XX
DR   WPI; 1996-321553/32.
XX
PT   Proliferating cell nuclear antigen binding substances, esp. fragment of
PT   p21WAF1  - used for treating disorders in which PCNA is implicated,
PT   partic. hyper-proliferative disorders, e.g. cancer or psoriasis
XX
PS   Disclosure; Fig 8; 35pp; English.
XX
CC   p21WAF1 is a protein that may be transcriptionally induced by the tumour
CC   suppressor p53, and acts as a potent inhibitor of cyclin dependent
CC   kinases (Cdks) in G1 and S phases of the cell cycle. p21WAF1 also binds
CC   to proliferating cell nuclear antigen (PCNA) at high concn. in vitro and
CC   blocks DNA replication. PCNA is a processivity factor for polymerase
CC   delta which plays an essential role in DNA replication and repair.
CC   During p53-mediated suppression of cell proliferation, p21WAF1 is
CC   important for co-ordinating cell cycle progression, DNA replication and
CC   repair of damaged DNA. In partic. peptides derived from the C-terminal
CC   region of p21WAF1 bind to PCNA and this accounts for the inhibition of
CC   DNA replication. The interaction of p21WAF1 with cyclin-Cdks and PCNA
CC   provides the possibility of using p21WAF1 to co-ordinate cell
CC   proliferation and cell cycle control, and in partic. to be used in
CC   treatment of hyperproliferative diseases such as cancer or psoriasis.
CC   Alanine scanning of p21WAF1 peptide 10 (AAR98310) was carried out to
CC   determine the PCNA binding capacity.  Each amino acid in the putative
CC   PCNA binding site was sequentially changed to alanine ("44"-"63") and in
CC   peptide 64 two arginines were altered to alanine within the same
CC   peptide.
XX
SQ   Sequence   20 AA;

  Query Match             100.0%;  Score 41;  DB 17;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 0.12;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          ||||||||
Db     12 HAKRRLIF 19


RESULT 6
```

```
AAU05708
ID   AAU05708 standard; Protein; 8 AA.
XX
AC   AAU05708;
XX
DT   21-NOV-2001  (first entry)
XX
DE   p21 C-terminus derived peptide #75.
XX
KW   Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW   inhibitor; proliferative disorder; cancer; leukaemia;
KW   drug screening.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "The N-terminus is hydrogenated"
FT   Modified-site   8
FT                   /note= "C-terminal amide"
XX
PN   WO200140142-A2.
XX
PD   07-JUN-2001.
XX
PF   29-NOV-2000; 2000WO-GB04550.
XX
PR   30-NOV-1999;   99GB-0028323.
XX
PA   (CYCL-) CYCLACEL LTD.
XX
PI   Zheleva DI,  Fischer PM,  McInnes C,  Andrews MJI,  Chan WC;
PI   Atkinson GE;
XX
DR   WPI; 2001-488493/53.
XX
PT   New p21 derived peptides and their variants, particularly useful as
PT   selective inhibitors of CDK2/cyclin interaction for treating
PT   proliferative disorders e.g. cancers and leukaemias, and in assays for
PT   identifying CDK/cyclin inhibitors -
XX
PS   Claim 25; Page 88; 102pp; English.
XX
CC   The invention relates to peptide and their variants derived from p21WAFI,
CC   which are inhibitors of CDK2 activity by binding to G1 and
CC   S phase specific cyclins which activate CDK2; selective inhibitors of
CC   CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
CC   The variants of the peptide may have further amino acids at either end
CC   or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC   The peptides are specific regions of p21WAFI that bind to G1 and S
CC   phase specific cyclins, preferably cyclins which activate CDK2. One
CC   of the peptides corresponds to p21(149-159). The peptides are used for
CC   treating proliferative disorders, e.g. cancers and leukaemias. The
CC   peptides are also for identifying substances which interfere with
CC   protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC   especially CDK/cyclin interactions, and which are capable of inhibiting
CC   CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC   competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC   may be used to identify substances that bind to, or inhibit peptide-
CC   cyclin interactions. Substances for screening in the assays include
CC   antibody products specific for p21 or cyclin binding regions,
CC   combinatorial libraries and single compound collections. The present
CC   sequence is a peptide derived from the C-terminus of p21.
XX
SQ   Sequence   8 AA;

  Query Match             97.6%;  Score 40;  DB 22;  Length 8;
  Best Local Similarity   87.5%;  Pred. No. 7.8e+05;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          ||||||:|
```

```
Db      1 HAKRRLVF 8


RESULT 7
AAU05702
ID   AAU05702 standard; Protein; 8 AA.
XX
AC   AAU05702;
XX
DT   21-NOV-2001  (first entry)
XX
DE   p21 C-terminus derived peptide #69.
XX
KW   Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW   inhibitor; proliferative disorder; cancer; leukaemia;
KW   drug screening.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "The N-terminus is hydrogenated"
FT   Modified-site   8
FT                   /note= "C-terminal amide"
XX
PN   WO200140142-A2.
XX
PD   07-JUN-2001.
XX
PF   29-NOV-2000; 2000WO-GB04550.
XX
PR   30-NOV-1999;   99GB-0028323.
XX
PA   (CYCL-) CYCLACEL LTD.
XX
PI   Zheleva DI,  Fischer PM,  McInnes C,  Andrews MJI,  Chan WC;
PI   Atkinson GE;
XX
DR   WPI; 2001-488493/53.
XX
PT   New p21 derived peptides and their variants, particularly useful as
PT   selective inhibitors of CDK2/cyclin interaction for treating
PT   proliferative disorders e.g. cancers and leukaemias, and in assays for
PT   identifying CDK/cyclin inhibitors -
XX
PS   Claim 25; Page 88; 102pp; English.
XX
CC   The invention relates to peptide and their variants derived from p21WAFI,
CC   which are inhibitors of CDK2 activity by binding to G1 and
CC   S phase specific cyclins which activate CDK2; selective inhibitors of
CC   CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
CC   The variants of the peptide may have further amino acids at either end
CC   or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC   The peptides are specific regions of p21WAFI that bind to G1 and S
CC   phase specific cyclins, preferably cyclins which activate CDK2. One
CC   of the peptides corresponds to p21(149-159). The peptides are used for
CC   treating proliferative disorders, e.g. cancers and leukaemias. The
CC   peptides are also for identifying substances which interfere with
CC   protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC   especially CDK/cyclin interactions, and which are capable of inhibiting
CC   CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC   competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC   may be used to identify substances that bind to, or inhibit peptide-
CC   cyclin interactions. Substances for screening in the assays include
CC   antibody products specific for p21 or cyclin binding regions,
CC   combinatorial libraries and single compound collections. The present
CC   sequence is a peptide derived from the C-terminus of p21.
XX
SQ   Sequence   8 AA;

  Query Match             95.1%;  Score 39;  DB 22;  Length 8;
  Best Local Similarity   87.5%;  Pred. No. 7.8e+05;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          |||||:||
Db      1 HAKRRIIF 8


RESULT 8
AAU05707
ID   AAU05707 standard; Protein; 8 AA.
XX
AC   AAU05707;
XX
DT   21-NOV-2001  (first entry)
XX
DE   p21 C-terminus derived peptide #74.
XX
KW   Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW   inhibitor; proliferative disorder; cancer; leukaemia;
KW   drug screening.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "The N-terminus is hydrogenated"
FT   Modified-site   8
FT                   /note= "C-terminal amide"
XX
PN   WO200140142-A2.
XX
PD   07-JUN-2001.
XX
PF   29-NOV-2000; 2000WO-GB04550.
XX
PR   30-NOV-1999;   99GB-0028323.
XX
PA   (CYCL-) CYCLACEL LTD.
XX
PI   Zheleva DI,  Fischer PM,  McInnes C,  Andrews MJI,  Chan WC;
PI   Atkinson GE;
XX
DR   WPI; 2001-488493/53.
XX
PT   New p21 derived peptides and their variants, particularly useful as
PT   selective inhibitors of CDK2/cyclin interaction for treating
PT   proliferative disorders e.g. cancers and leukaemias, and in assays for
PT   identifying CDK/cyclin inhibitors -
XX
PS   Claim 25; Page 88; 102pp; English.
XX
CC   The invention relates to peptide and their variants derived from p21WAFI,
CC   which are inhibitors of CDK2 activity by binding to G1 and
CC   S phase specific cyclins which activate CDK2; selective inhibitors of
CC   CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
CC   The variants of the peptide may have further amino acids at either end
CC   or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC   The peptides are specific regions of p21WAFI that bind to G1 and S
CC   phase specific cyclins, preferably cyclins which activate CDK2. One
CC   of the peptides corresponds to p21(149-159). The peptides are used for
CC   treating proliferative disorders, e.g. cancers and leukaemias. The
CC   peptides are also for identifying substances which interfere with
CC   protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC   especially CDK/cyclin interactions, and which are capable of inhibiting
CC   CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC   competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC   may be used to identify substances that bind to, or inhibit peptide-
CC   cyclin interactions. Substances for screening in the assays include
CC   antibody products specific for p21 or cyclin binding regions,
CC   combinatorial libraries and single compound collections. The present
CC   sequence is a peptide derived from the C-terminus of p21.
XX
```

```
SQ   Sequence   8 AA;

  Query Match             95.1%;  Score 39;  DB 22;  Length 8;
  Best Local Similarity   87.5%;  Pred. No. 7.8e+05;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          ||||||:|
Db      1 HAKRRLLF 8


RESULT 9
AAU05742
ID   AAU05742 standard; Protein; 8 AA.
XX
AC   AAU05742;
XX
DT   21-NOV-2001  (first entry)
XX
DE   p21 C-terminus derived peptide #111.
XX
KW   Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW   inhibitor; proliferative disorder; cancer; leukaemia;
KW   drug screening.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "The N-terminus is hydrogenated"
FT   Modified-site   8
FT                   /note= "C-terminal amide"
XX
PN   WO200140142-A2.
XX
PD   07-JUN-2001.
XX
PF   29-NOV-2000; 2000WO-GB04550.
XX
PR   30-NOV-1999;   99GB-0028323.
XX
PA   (CYCL-) CYCLACEL LTD.
XX
PI   Zheleva DI,  Fischer PM,  McInnes C,  Andrews MJI,  Chan WC;
PI   Atkinson GE;
XX
DR   WPI; 2001-488493/53.
XX
PT   New p21 derived peptides and their variants, particularly useful as
PT   selective inhibitors of CDK2/cyclin interaction for treating
PT   proliferative disorders e.g. cancers and leukaemias, and in assays for
PT   identifying CDK/cyclin inhibitors -
XX
PS   Claim 34; Page 92; 102pp; English.
XX
CC   The invention relates to peptide and their variants derived from p21WAFI,
CC   which are inhibitors of CDK2 activity by binding to G1 and
CC   S phase specific cyclins which activate CDK2; selective inhibitors of
CC   CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
CC   The variants of the peptide may have further amino acids at either end
CC   or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC   The peptides are specific regions of p21WAFI that bind to G1 and S
CC   phase specific cyclins, preferably cyclins which activate CDK2. One
CC   of the peptides corresponds to p21(149-159). The peptides are used for
CC   treating proliferative disorders, e.g. cancers and leukaemias. The
CC   peptides are also for identifying substances which interfere with
CC   protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC   especially CDK/cyclin interactions, and which are capable of inhibiting
CC   CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC   competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC   may be used to identify substances that bind to, or inhibit peptide-
CC   cyclin interactions. Substances for screening in the assays include
CC   antibody products specific for p21 or cyclin binding regions,
CC   combinatorial libraries and single compound collections. The present
CC   sequence is a peptide derived from the C-terminus of p21.
XX
SQ   Sequence   8 AA;

  Query Match             95.1%;  Score 39;  DB 22;  Length 8;
  Best Local Similarity   87.5%;  Pred. No. 7.8e+05;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          ||||||:|
Db      1 HAKRRLLF 8


RESULT 10
AAG65127
ID   AAG65127 standard; Peptide; 8 AA.
XX
AC   AAG65127;
XX
DT   21-NOV-2001  (first entry)
XX
DE   p21WAFI C-terminal peptide #30.
XX
KW   Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW   inhibitor; proliferative disorder; cancer; leukaemia; drug screening.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   8
FT                   /note= "Optional C-terminal carboxamide"
XX
PN   WO200140142-A2.
XX
PD   07-JUN-2001.
XX
PF   29-NOV-2000; 2000WO-GB04550.
XX
PR   30-NOV-1999;   99GB-0028323.
XX
PA   (CYCL-) CYCLACEL LTD.
XX
PI   Zheleva DI,  Fischer PM,  McInnes C,  Andrews MJI,  Chan WC;
PI   Atkinson GE;
XX
DR   WPI; 2001-488493/53.
XX
PT   New p21 derived peptides and their variants, particularly useful as
PT   selective inhibitors of CDK2/cyclin interaction for treating
PT   proliferative disorders e.g. cancers and leukaemias, and in assays for
PT   identifying CDK/cyclin inhibitors -
XX
PS   Claim 15; Page 84; 102pp; English.
XX
CC   The invention relates to peptide and their variants derived from p21WAFI,
CC   which are inhibitors of CDK2 activity by binding to G1 and
CC   S phase specific cyclins which activate CDK2; selective inhibitors of
CC   CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
CC   The variants of the peptide may have further amino acids at either end
CC   or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC   The peptides are specific regions of p21WAFI that bind to G1 and S
CC   phase specific cyclins, preferably cyclins which activate CDK2. One
CC   of the peptides corresponds to p21(149-159). The peptides are used for
CC   treating proliferative disorders, e.g. cancers and leukaemias. The
CC   peptides are also for identifying substances which interfere with
CC   protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC   especially CDK/cyclin interactions, and which are capable of inhibiting
CC   CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC   competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC   may be used to identify substances that bind to, or inhibit peptide-
CC   cyclin interactions. Substances for screening in the assays include
```

```
CC   antibody products specific for p21 or cyclin binding regions,
CC   combinatorial libraries and single compound collections. The present
CC   sequence is a peptide corresponding to p21(145-164) or a peptide
CC   derived from that region.
XX
SQ   Sequence   8 AA;

  Query Match             92.7%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   87.5%;  Pred. No. 7.8e+05;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          |:||||||
Db      1 HSKRRLIF 8


RESULT 11
AAG66266
ID   AAG66266 standard; Peptide; 8 AA.
XX
AC   AAG66266;
XX
DT   21-NOV-2001  (first entry)
XX
DE   p21 C-terminus derived peptide #58.
XX
KW   Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW   inhibitor; proliferative disorder; cancer; leukaemia;
KW   drug screening.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "The N-terminus is hydrogenated"
FT   Modified-site   8
FT                   /note= "C-terminal amide"
XX
PN   WO200140142-A2.
XX
PD   07-JUN-2001.
XX
PF   29-NOV-2000; 2000WO-GB04550.
XX
PR   30-NOV-1999;   99GB-0028323.
XX
PA   (CYCL-) CYCLACEL LTD.
XX
PI   Zheleva DI,  Fischer PM,  McInnes C,  Andrews MJI,  Chan WC;
PI   Atkinson GE;
XX
DR   WPI; 2001-488493/53.
XX
PT   New p21 derived peptides and their variants, particularly useful as
PT   selective inhibitors of CDK2/cyclin interaction for treating
PT   proliferative disorders e.g. cancers and leukaemias, and in assays for
PT   identifying CDK/cyclin inhibitors -
XX
PS   Claim 25; Page 87; 102pp; English.
XX
CC   The invention relates to peptide and their variants derived from p21WAFI,
CC   which are inhibitors of CDK2 activity by binding to G1 and
CC   S phase specific cyclins which activate CDK2; selective inhibitors of
CC   CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
CC   The variants of the peptide may have further amino acids at either end
CC   or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC   The peptides are specific regions of p21WAFI that bind to G1 and S
CC   phase specific cyclins, preferably cyclins which activate CDK2. One
CC   of the peptides corresponds to p21(149-159). The peptides are used for
CC   treating proliferative disorders, e.g. cancers and leukaemias. The
CC   peptides are also for identifying substances which interfere with
CC   protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC   especially CDK/cyclin interactions, and which are capable of inhibiting
CC   CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC   competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC   may be used to identify substances that bind to, or inhibit peptide-
CC   cyclin interactions. Substances for screening in the assays include
CC   antibody products specific for p21 or cyclin binding regions,
CC   combinatorial libraries and single compound collections. The present
CC   sequence is a peptide derived from the C-terminus of p21.
XX
SQ   Sequence   8 AA;

  Query Match             92.7%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   87.5%;  Pred. No. 7.8e+05;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          ||:|||||
Db      1 HARRRLIF 8


RESULT 12
AAU05703
ID   AAU05703 standard; Protein; 8 AA.
XX
AC   AAU05703;
XX
DT   21-NOV-2001  (first entry)
XX
DE   p21 C-terminus derived peptide #70.
XX
KW   Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW   inhibitor; proliferative disorder; cancer; leukaemia;
KW   drug screening.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "The N-terminus is hydrogenated"
FT   Modified-site   8
FT                   /note= "C-terminal amide"
XX
PN   WO200140142-A2.
XX
PD   07-JUN-2001.
XX
PF   29-NOV-2000; 2000WO-GB04550.
XX
PR   30-NOV-1999;   99GB-0028323.
XX
PA   (CYCL-) CYCLACEL LTD.
XX
PI   Zheleva DI,  Fischer PM,  McInnes C,  Andrews MJI,  Chan WC;
PI   Atkinson GE;
XX
DR   WPI; 2001-488493/53.
XX
PT   New p21 derived peptides and their variants, particularly useful as
PT   selective inhibitors of CDK2/cyclin interaction for treating
PT   proliferative disorders e.g. cancers and leukaemias, and in assays for
PT   identifying CDK/cyclin inhibitors -
XX
PS   Claim 25; Page 88; 102pp; English.
XX
CC   The invention relates to peptide and their variants derived from p21WAFI,
CC   which are inhibitors of CDK2 activity by binding to G1 and
CC   S phase specific cyclins which activate CDK2; selective inhibitors of
CC   CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
CC   The variants of the peptide may have further amino acids at either end
CC   or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC   The peptides are specific regions of p21WAFI that bind to G1 and S
CC   phase specific cyclins, preferably cyclins which activate CDK2. One
```

```
CC   of the peptides corresponds to p21(149-159). The peptides are used for
CC   treating proliferative disorders, e.g. cancers and leukaemias. The
CC   peptides are also for identifying substances which interfere with
CC   protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC   especially CDK/cyclin interactions, and which are capable of inhibiting
CC   CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC   competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC   may be used to identify substances that bind to, or inhibit peptide-
CC   cyclin interactions. Substances for screening in the assays include
CC   antibody products specific for p21 or cyclin binding regions,
CC   combinatorial libraries and single compound collections. The present
CC   sequence is a peptide derived from the C-terminus of p21.
XX
SQ   Sequence   8 AA;

  Query Match             92.7%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   87.5%;  Pred. No. 7.8e+05;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          |||||:||
Db      1 HAKRRVIF 8


RESULT 13
AAU05714
ID   AAU05714 standard; Protein; 8 AA.
XX
AC   AAU05714;
XX
DT   21-NOV-2001  (first entry)
XX
DE   p21 C-terminus derived peptide #81.
XX
KW   Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW   inhibitor; proliferative disorder; cancer; leukaemia;
KW   drug screening.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "The N-terminus is hydrogenated"
FT   Modified-site   8
FT                   /note= "C-terminal amide"
XX
PN   WO200140142-A2.
XX
PD   07-JUN-2001.
XX
PF   29-NOV-2000; 2000WO-GB04550.
XX
PR   30-NOV-1999;   99GB-0028323.
XX
PA   (CYCL-) CYCLACEL LTD.
XX
PI   Zheleva DI,  Fischer PM,  McInnes C,  Andrews MJI,  Chan WC;
PI   Atkinson GE;
XX
DR   WPI; 2001-488493/53.
XX
PT   New p21 derived peptides and their variants, particularly useful as
PT   selective inhibitors of CDK2/cyclin interaction for treating
PT   proliferative disorders e.g. cancers and leukaemias, and in assays for
PT   identifying CDK/cyclin inhibitors -
XX
PS   Claim 25; Page 88; 102pp; English.
XX
CC   The invention relates to peptide and their variants derived from p21WAFI,
CC   which are inhibitors of CDK2 activity by binding to G1 and
CC   S phase specific cyclins which activate CDK2; selective inhibitors of
CC   CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
CC   The variants of the peptide may have further amino acids at either end
CC   or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC   The peptides are specific regions of p21WAFI that bind to G1 and S
CC   phase specific cyclins, preferably cyclins which activate CDK2. One
CC   of the peptides corresponds to p21(149-159). The peptides are used for
CC   treating proliferative disorders, e.g. cancers and leukaemias. The
CC   peptides are also for identifying substances which interfere with
CC   protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC   especially CDK/cyclin interactions, and which are capable of inhibiting
CC   CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC   competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC   may be used to identify substances that bind to, or inhibit peptide-
CC   cyclin interactions. Substances for screening in the assays include
CC   antibody products specific for p21 or cyclin binding regions,
CC   combinatorial libraries and single compound collections. The present
CC   sequence is a peptide derived from the C-terminus of p21.
XX
SQ   Sequence   8 AA;

  Query Match             92.7%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   87.5%;  Pred. No. 7.8e+05;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 HAKRRLIF 8
          |||||||:
Db      1 HAKRRLIY 8


RESULT 14
AAG65110
ID   AAG65110 standard; Peptide; 9 AA.
XX
AC   AAG65110;
XX
DT   21-NOV-2001  (first entry)
XX
DE   p21WAFI C-terminal peptide #13.
XX
KW   Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW   inhibitor; proliferative disorder; cancer; leukaemia; drug screening.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   9
FT                   /note= "Optional C-terminal carboxamide"
XX
PN   WO200140142-A2.
XX
PD   07-JUN-2001.
XX
PF   29-NOV-2000; 2000WO-GB04550.
XX
PR   30-NOV-1999;   99GB-0028323.
XX
PA   (CYCL-) CYCLACEL LTD.
XX
PI   Zheleva DI,  Fischer PM,  McInnes C,  Andrews MJI,  Chan WC;
PI   Atkinson GE;
XX
DR   WPI; 2001-488493/53.
XX
PT   New p21 derived peptides and their variants, particularly useful as
PT   selective inhibitors of CDK2/cyclin interaction for treating
PT   proliferative disorders e.g. cancers and leukaemias, and in assays for
PT   identifying CDK/cyclin inhibitors -
XX
PS   Claim 15; Page 84; 102pp; English.
XX
CC   The invention relates to peptide and their variants derived from p21WAFI,
CC   which are inhibitors of CDK2 activity by binding to G1 and
CC   S phase specific cyclins which activate CDK2; selective inhibitors of
CC   CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
```

CC The variants of the peptide may have further amino acids at either end
CC or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC The peptides are specific regions of p21WAFI that bind to G1 and S
CC phase specific cyclins, preferably cyclins which activate CDK2. One
CC of the peptides corresponds to p21(149-159). The peptides are used for
CC treating proliferative disorders, e.g. cancers and leukaemias. The
CC peptides are also for identifying substances which interfere with
CC protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC especially CDK/cyclin interactions, and which are capable of inhibiting
CC CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC may be used to identify substances that bind to, or inhibit peptide-
CC cyclin interactions. Substances for screening in the assays include
CC antibody products specific for p21 or cyclin binding regions,
CC combinatorial libraries and single compound collections. The present
CC sequence is a peptide corresponding to p21(145-164) or a peptide
CC derived from that region.
XX
SQ Sequence 9 AA;

Query Match 92.7%; Score 38; DB 22; Length 9;
Best Local Similarity 87.5%; Pred. No. 7.8e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HAKRRLIF 8
|:||||||
Db 1 HSKRRLIF 8

RESULT 15
AAG65122
ID AAG65122 standard; Peptide; 9 AA.
XX
AC AAG65122;
XX
DT 21-NOV-2001 (first entry)
XX
DE p21WAFI C-terminal peptide #25.
XX
KW Human; p21WAFI; cyclin dependent protein kinase; CDK2; cyclin A;
KW inhibitor; proliferative disorder; cancer; leukaemia; drug screening.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 9
FT /note= "Optional C-terminal carboxamide"
XX
PN WO200140142-A2.
XX
PD 07-JUN-2001.
XX
PF 29-NOV-2000; 2000WO-GB04550.
XX
PR 30-NOV-1999; 99GB-0028323.
XX
PA (CYCL-) CYCLACEL LTD.
XX
PI Zheleva DI, Fischer PM, McInnes C, Andrews MJI, Chan WC;
PI Atkinson GE;
XX
DR WPI; 2001-488493/53.
XX
PT New p21 derived peptides and their variants, particularly useful as
PT selective inhibitors of CDK2/cyclin interaction for treating
PT proliferative disorders e.g. cancers and leukaemias, and in assays for
PT identifying CDK/cyclin inhibitors -
XX
PS Claim 15; Page 84; 102pp; English.
XX
CC The invention relates to peptide and their variants derived from p21WAFI,
CC which are inhibitors of CDK2 activity by binding to G1 and
CC S phase specific cyclins which activate CDK2; selective inhibitors of
CC CDK2/cyclin complexes, particularly CDK2/cyclin A or E complexes.
CC The variants of the peptide may have further amino acids at either end
CC or have up to 7 amino acids deleted, provided the motif XLXF is retained.
CC The peptides are specific regions of p21WAFI that bind to G1 and S
CC phase specific cyclins, preferably cyclins which activate CDK2. One
CC of the peptides corresponds to p21(149-159). The peptides are used for
CC treating proliferative disorders, e.g. cancers and leukaemias. The
CC peptides are also for identifying substances which interfere with
CC protein-protein interactions involving cyclins (i.e. cyclin A, E or D),
CC especially CDK/cyclin interactions, and which are capable of inhibiting
CC CDK2 and/or CDK4 activity. P21 peptides other than p21(149-159)
CC competitively inhibit the binding of peptide p21(149-159) to cyclin and
CC may be used to identify substances that bind to, or inhibit peptide-
CC cyclin interactions. Substances for screening in the assays include
CC antibody products specific for p21 or cyclin binding regions,
CC combinatorial libraries and single compound collections. The present
CC sequence is a peptide corresponding to p21(145-164) or a peptide
CC derived from that region.
XX
SQ Sequence 9 AA;

Query Match 92.7%; Score 38; DB 22; Length 9;
Best Local Similarity 87.5%; Pred. No. 7.8e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HAKRRLIF 8
|:||||||
Db 2 HSKRRLIF 9

Search completed: December 14, 2002, 15:45:43
Job time : 58 secs

GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2002, 15:50:04 ; Search time 220 Seconds
(without alignments)
81.891 Million cell updates/sec

Title: US-09-726-470A-35
Perfect score: 41
Sequence: 1 HAKRRLIF 8

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09726470/runat_10122002_090717_4962/app_query.fasta_1.398
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09726470_@CGN_1_1_79_@runat_10122002_090717_4962 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	38	92.7	80	21	AAZ30402	PCR primer used to
	2	38	92.7	331	22	AAF79981	Nucleotide sequenc
	3	38	92.7	495	19	AAV16718	Nucleotide sequenc
	4	38	92.7	1194	16	AAQ86776	GST-SDI-1 gene fus
	5	38	92.7	2106	14	AAQ43905	Sequence encoding
	6	38	92.7	2106	16	AAQ86773	SDI-1 cDNA. Homo
	7	38	92.7	2106	17	AAT18792	Senescent cell-der
	8	38	92.7	2106	17	AAT06940	Senescent cell der
	9	38	92.7	2121	16	AAQ90445	Human WAF1 gene cD
	10	38	92.7	2121	18	AAT61419	Human WAF1 cDNA.
	11	38	92.7	2121	20	AAX15105	cDNA encoding a pr
	12	38	92.7	2121	24	ABK84187	Human cDNA differe
	13	38	92.7	2121	24	AAI72397	p21-Cip1 cDNA. Ho
	14	38	92.7	2127	24	AAS94878	Human DNA sequence
	15	38	92.7	2147	16	AAQ90051	Melanoma different
	16	38	92.7	2342	21	AAC99083	Human pancreatic c
	17	37	90.2	65	24	ABN30342	Rat spliced transc
	18	36	87.8	18189	23	ABL16916	Drosophila melanog
c	19	35	85.4	240	21	AAC93606	Cat flea head and
c	20	35	85.4	56743	22	AAK68202	Human immune/haema
c	21	35	85.4	56743	22	AAK81760	Human immune/haema
	22	34	82.9	91	21	AAC32321	Human secreted pro
	23	34	82.9	102	21	AAC32492	Human secreted pro
c	24	34	82.9	172	24	ABL85312	Human ovarian canc
c	25	34	82.9	387	24	ABN90790	Staphylococcus epi
	26	34	82.9	407	23	ABV17984	Human prostate exp
	27	34	82.9	421	23	AAS77290	DNA encoding novel
	28	34	82.9	476	23	ABV47773	Human prostate exp
c	29	34	82.9	568	22	AAK37833	Human bone marrow
c	30	34	82.9	568	24	ABS11827	Human genome-deriv
c	31	34	82.9	819	14	AAQ38089	19857 Osp A DNA.
c	32	34	82.9	1014	14	AAQ38091	NS1-19857 fusion c
c	33	34	82.9	1108	24	ABQ49326	Oligonucleotide fo
	34	34	82.9	1108	24	ABQ49327	Oligonucleotide fo
c	35	34	82.9	1111	24	ABQ45652	Oligonucleotide fo
	36	34	82.9	1111	24	ABQ45653	Oligonucleotide fo
c	37	34	82.9	2248	21	AAC59806	Human secreted pro
c	38	34	82.9	2903	21	AAA71628	Human aspartate pr
	39	34	82.9	4047	22	AAH54462	S. epidermidis gen
c	40	34	82.9	4290	23	ABL52917	2-keto-D-gluconate
c	41	34	82.9	4444	23	ABL05116	Drosophila melanog
c	42	34	82.9	14822	20	AAX20543	Polynucleotide seq
	43	34	82.9	32247	22	ABA19669	Human nervous syst
	44	33	80.5	88	22	AAI27907	Probe #17840 for g
	45	33	80.5	88	22	AAI56890	Probe #25576 used

ALIGNMENTS

RESULT 1
AAZ30402
ID AAZ30402 standard; DNA; 80 BP.
XX
AC AAZ30402;
XX
DT 11-FEB-2000 (first entry)
XX
DE PCR primer used to amplify a fragment of p21 gene cDNA.
XX
KW Bioactive agent; cellular phenotype; fluorescence activated cell sorting;
KW FACS; exocytosis; allergy; asthma; rhinitis; psychiatric disorder;
KW Chediak-Higashi syndrome; mouse; mink; cattle; cat; killer whale;
KW cell cycle regulation; cancer; p21; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX

PN WO9954494-A2.
XX
PD 28-OCT-1999.
XX
PF 16-APR-1999; 99WO-US08345.
XX
PR 17-APR-1998; 98US-0062330.
PR 21-SEP-1998; 98US-0157748.
XX
PA (RIGE-) RIGEL PHARM INC.
XX
PI Fisher J, Lorens J, Payan D, Rossi A;
XX
DR WPI; 2000-013265/01.
XX
PT New method for screening for agents which alter a cellular phenotype,
PT used for identifying agents for treating e.g. tumours, allergy, asthma
PT or psychiatric disorders -
XX
PS Example 1; Page 45; 69pp; English.
XX
CC The specification describes a method of screening for a bioactive agent
CC capable of altering a cellular phenotype. The method comprises combining
CC at least one candidate bioactive agent and a population of cells; and
CC sorting the cells in a fluorescence activated cell sorting (FACS)
CC machine by separating the cells on the basis of at least 5 cellular
CC parameters. The methods can be used for identifying agents for treating
CC disorders involving exocytosis, e.g. allergy, asthma, rhinitis,
CC psychiatric disorders or Chediak-Higashi syndrome and similar disorders
CC in mice, mink, cattle, cats, and killer whales. They can also be used
CC for identifying agents for treating disorders involving cell cycle
CC regulation such as cancers. They can also be used for identifying agents
CC which alter other cellular phenotypes, e.g. small molecule toxicity or
CC the expression of moieties e.g. receptors (particularly cell surface
CC receptors), adhesion molecules, cytokine secretion, or protein-protein
CC interactions. PCR primers AAZ30402-03 were used to amplify a fragment
CC of p21 cDNA. The amplified sequence was used in the method of the
CC invention, in constructs for cell cycle assays using p21 as a positive
CC control.
XX
SQ Sequence 80 BP; 22 A; 29 C; 16 G; 13 T; 0 other;

Alignment Scores:

Pred. No.:	5.05	Length:	80
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	92.68%	Indels:	0
DB:	21	Gaps:	0

US-09-726-470A-35 (1-8) x AAZ30402 (1-80)

```
Qy       1 HisAlaLysArgArgLeuIlePhe 8
           |||:::|||||||||||||||||||
Db      52 CACTCCAAACGCCGGCTGATCTTC 75
```

RESULT 2
AAF79981
ID AAF79981 standard; DNA; 331 BP.
XX
AC AAF79981;
XX
DT 11-JUN-2001 (first entry)
XX
DE Nucleotide sequence of a human genetic marker for toxicity.
XX
KW Genetic marker; toxicity; cellular signalling pathway; polymorphism; ss.
XX
OS Homo sapiens.
XX
PN WO200120029-A2.
XX
PD 22-MAR-2001.
XX
PF 12-SEP-2000; 2000WO-FR02503.
XX
PR 13-SEP-1999; 99FR-0011405.
XX
PA (EXON-) EXONHIT THERAPEUTICS SA.
XX
PI Tocque B, Bracco L, Schweighoffer F;
XX
DR WPI; 2001-244821/25.
XX
PT Analysing the toxic potential of test compounds, for e.g. screening
PT for toxic effects in potential pharmaceuticals, comprises analysing
PT hybridisation patterns of treated cells -
XX
PS Claim 35; Page 61; 68pp; French.
XX
CC AAF79967-AAF80003 represents genetic markers of toxicity. The
CC specification describes a method for analysing the toxic potential
CC of a test compound. The method comprises hybridising nucleic acids
CC from cells treated with the test compound and the present markers.
CC These markers correspond to genetic events characteristic of
CC deregulation of cellular signalling pathways. The method is used to
CC identify the toxic potential of compounds (particularly human or
CC veterinary pharmaceuticals or plant protection agents) and to evaluate
CC the response and/or sensitivity of subjects to a particular compound,
CC from the presence of polymorphisms or other mutations in particular
CC genes.
XX
SQ Sequence 331 BP; 81 A; 98 C; 63 G; 84 T; 5 other;

Alignment Scores:

Pred. No.:	24.8	Length:	331
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	92.68%	Indels:	0
DB:	22	Gaps:	0

US-09-726-470A-35 (1-8) x AAF79981 (1-331)

```
Qy       1 HisAlaLysArgArgLeuIlePhe 8
           |||:::|||||||||||||||||||
Db      33 CACTCCAAACGCCGGTTGATCTTC 56
```

RESULT 3
AAV16718
ID AAV16718 standard; cDNA; 495 BP.
XX
AC AAV16718;
XX
DT 15-JUN-1998 (first entry)
XX
DE Nucleotide sequence encoding the p21CIP1 protein.
XX
KW E7 oncoprotein; proliferative state; HPV; kinase activity;
KW cyclin/cyclin-dependent kinase; p21CIP1; interaction; inactivation;
KW cyclin/cyclin-dependent kinase inhibitor; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..495
FT /*tag= a
XX
PN US5736318-A.
XX
PD 07-APR-1998.
XX
PF 17-MAR-1995; 95US-0406248.
XX
PR 17-MAR-1995; 95US-0406248.
XX

PA (HARD) HARVARD COLLEGE.
PA (HARD) UNIV HARVARD.
XX
PI Jones DL, Munger K;
XX
DR WPI; 1998-239202/21.
DR P-PSDB; AAW46887.
XX
PT Evaluation of proliferative state of cells transformed with human
PT papilloma virus - by determining cyclin-dependent kinase activity
PT induced by E7 onco-protein
XX
PS Disclosure; Columns 11-14; 14pp; English.
XX
CC The present sequence encodes a p21CIP1 protein, which is part of a
CC family of small cyclin-dependent kinase inhibitors. The proliferative
CC state of a cell transformed with Human papillomavirus (HPV) can be
CC evaluated in the following manner. Cyclin/cyclin-dependent kinase
CC complexes containing protein p21CIP1 are isolated from the transformed
CC cell, and the HPV E7 oncoprotein (AAW46886) added to the isolated
CC protein. Cyclin/cyclin-dependent kinase complexes are isolated from an
CC untransformed cell that is substantially homogenic with the transformed
CC cell, and the HPV E7 oncoprotein added. The kinase activities of the 2
CC samples are measured, where a proliferating transformed cell has a
CC greater kinase activity that the untransformed cell. The method is
CC used for determining the extent of interaction and/or inactivation
CC between a cyclin/cyclin-dependent kinase inhibitor and the HPV E7
CC oncoprotein and thus evaluating the proliferative state of a transformed
CC cell.
XX
SQ Sequence 495 BP; 96 A; 150 C; 165 G; 84 T; 0 other;

Alignment Scores:

Pred. No.:	39	Length:	495
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	92.68%	Indels:	0
DB:	19	Gaps:	0

US-09-726-470A-35 (1-8) x AAV16718 (1-495)

```
Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db    454 CACTCCAAACGCCGGCTGATCTTC 477
```

RESULT 4
AAQ86776
ID AAQ86776 standard; cDNA; 1194 BP.
XX
AC AAQ86776;
XX
DT 16-OCT-1995 (first entry)
XX
DE GST-SDI-1 gene fusion.
XX
KW SDI-1; senescent cell-derived inhibitor; DNA synthesis;
KW senescence; cell proliferation; cancer; therapeutic; vulnerary;
KW fusion protein; glutathione-s-transferase; ss.
XX
OS Synthetic.
XX
PN WO9506415-A.
XX
PD 09-MAR-1995.
XX
PF 26-AUG-1994; 94WO-US09700.
XX
PR 13-JUL-1994; 94US-0274535.
PR 30-AUG-1993; 93US-0113372.
PR 17-NOV-1993; 93US-0153564.
PR 03-JAN-1994; 94US-0160814.
PR 25-FEB-1994; 94US-0203535.
PR 15-APR-1994; 94US-0229420.
PR 30-JUN-1994; 94US-0268439.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Smith JR;
XX
DR WPI; 1995-131101/17.
DR P-PSDB; AAR72795.
XX
PT Nucleic acid encoding a protein or polypeptide that inhibits DNA
PT synthesis in a recipient cell - useful to inhibit cell
PT proliferation in tumour cells, treat wound or burn tissue, or as
PT an antiviral or antiparasitic agent
XX
PS Disclosure; Page 133; 169pp; English.
XX
CC The gene fusion sequence given in AAQ86776 encodes amino acids 1-226
CC of Schistosoma japonicum glutathione-s-transferase fused to a
CC 9-amino acid hinge region (AAR86775) and the human senescent
CC cell-derived inhibitor SDI-1 (AAR72791). The resulting fusion
CC protein (AAR72795) was expressed in E. coli transformed with
CC plasmid pAG20 (ATCC 69597).
XX
SQ Sequence 1194 BP; 299 A; 278 C; 321 G; 296 T; 0 other;

Alignment Scores:

Pred. No.:	105	Length:	1194
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	92.68%	Indels:	0
DB:	16	Gaps:	0

US-09-726-470A-35 (1-8) x AAQ86776 (1-1194)

```
Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db   1153 CACTCCAAACGCCGGCTGATCTTC 1176
```

RESULT 5
AAQ43905
ID AAQ43905 standard; cDNA; 2106 BP.
XX
AC AAQ43905;
XX
DT 01-DEC-1993 (first entry)
XX
DE Sequence encoding a putative senescent cell derived inhibitor (SDI-1).
XX
KW Senescent cell derived inhibitor; SDI-1; infertility;
KW vascularisation; tissue regeneration; cancer therapy; wound-healing;
KW ageing; gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 79..574
FT /*tag= a
XX
PN WO9312251-A.
XX
PD 24-JUN-1993.
XX
PF 15-DEC-1992; 92WO-US10904.
XX
PR 16-DEC-1991; 91US-0808523.
PR 02-NOV-1992; 92US-0970462.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Smith JR;
XX

```
DR   WPI; 1993-214192/26.
DR   P-PSDB; AAR38299.
XX
PT   Senescent cell derived inhibitors - used to treat cancer and for
PT   wound healing and anti-ageing by antisense mechanisms
XX
PS   Claim 4; Fig 5; 57pp; English.
XX
CC   Efficient DEAE dextran-mediated transfection enabled the isolation
CC   of putative SDI sequences in three distinct cDNA clones. The
CC   expression of one, SDI-1, increased 20-fold at cellular senescence,
CC   whereas that of the other two remained constant. SDI-1 appears to be
CC   closely related to cellular sensescence. In the invention SDI-1 cDNA
CC   is incorporated into a DNA plasmid (such as pcDSR-alpha-delta).A
CC   plasid contg. SDI-1 was microinjected into the nuclei of young
CC   cycling cells. The plasmid showed a strong inhibitory action on the
CC   DNA synthesis of young cells with % inhibition of 57.5-62.4.
XX
SQ   Sequence 2106 BP; 404 A; 632 C; 575 G; 495 T; 0 other;

Alignment Scores:
Pred. No.:              198          Length:         2106
Score:                  38.00        Matches:        7
Percent Similarity:     100.00%      Conservative:   1
Best Local Similarity:  87.50%       Mismatches:     0
Query Match:            92.68%       Indels:         0
DB:                     14           Gaps:           0

US-09-726-470A-35 (1-8) x AAQ43905 (1-2106)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    532 CACTCCAAACGCCGGCTGATCTTC 555

RESULT 6
AAQ86773
ID   AAQ86773 standard; cDNA; 2106 BP.
XX
AC   AAQ86773;
XX
DT   16-OCT-1995  (first entry)
XX
DE   SDI-1 cDNA.
XX
KW   SDI-1; senescent cell-derived inhibitor; DNA synthesis;
KW   senescence; cell proliferation; cancer; therapeutic; vulnerary; ss.
XX
OS   Homo sapiens.
XX
PN   WO9506415-A.
XX
PD   09-MAR-1995.
XX
PF   26-AUG-1994;   94WO-US09700.
XX
PR   13-JUL-1994;   94US-0274535.
PR   30-AUG-1993;   93US-0113372.
PR   17-NOV-1993;   93US-0153564.
PR   03-JAN-1994;   94US-0160814.
PR   25-FEB-1994;   94US-0203535.
PR   15-APR-1994;   94US-0229420.
PR   30-JUN-1994;   94US-0268439.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Smith JR;
XX
DR   WPI; 1995-131101/17.
DR   P-PSDB; AAR72791.
XX
PT   Nucleic acid encoding a protein or polypeptide that inhibits DNA
PT   synthesis in a recipient cell  - useful to inhibit cell
PT   proliferation in tumour cells, treat wound or burn tissue, or as
PT   an antiviral or antiparasitic agent
XX
PS   Disclosure; Page 126-127; 169pp; English.
XX
CC   The gene encoding an inhibitor of DNA synthesis was identified by
CC   incorporating a senescent fibroblast cell cDNA library into a
CC   mammalian expression vector.  The library was then transfected into
CC   young, cycling cells to identify those library members that suppressed
CC   the inhibition of DNA synthesis.  DEAE dextran-mediated transfection
CC   enabled the isolation of senescent cell-derived inhibitor-1 (SDI-1),
CC   whose expression increased 20-fold at cellular senescence.
XX
SQ   Sequence 2106 BP; 404 A; 632 C; 575 G; 495 T; 0 other;

Alignment Scores:
Pred. No.:              198          Length:         2106
Score:                  38.00        Matches:        7
Percent Similarity:     100.00%      Conservative:   1
Best Local Similarity:  87.50%       Mismatches:     0
Query Match:            92.68%       Indels:         0
DB:                     16           Gaps:           0

US-09-726-470A-35 (1-8) x AAQ86773 (1-2106)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    532 CACTCCAAACGCCGGCTGATCTTC 555

RESULT 7
AAT18792
ID   AAT18792 standard; cDNA; 2106 BP.
XX
AC   AAT18792;
XX
DT   17-SEP-1996  (first entry)
XX
DE   Senescent cell-derived inhibitor cDNA clone SDI-1.
XX
KW   Senescent cell-derived inhibitor; SDI-1; DNA synthesis inhibitor;
KW   gene therapy; liposome; cancer; glaucoma; anti-ageing;
KW   senescence; cell proliferation; cancer; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             79..573
FT                   /*tag=  a
FT   conflict        286
FT                   /*tag=  b
FT                   /note= "G at position 286 differs from a previously
FT                          detd. sequence"
FT   conflict        1843..1844
FT                   /*tag=  c
FT                   /note= "CG at positions 1843-1844 differ from a
FT                          previously detd. sequence"
XX
PN   WO9612506-A1.
XX
PD   02-MAY-1996.
XX
PF   24-OCT-1995;   95WO-US13766.
XX
PR   06-SEP-1995;   95US-0524218.
PR   24-OCT-1994;   94US-0327874.
PR   23-MAY-1995;   95WO-US06451.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
PA   (SENN-) SENNES DRUG INNOVATIONS INC.
XX
PI   Drutz DG,  Smith JR,  Wilson DR,  Zumstein LA;
XX
DR   WPI; 1996-230371/23.
DR   P-PSDB; AAR94932.
```

```
XX
PT   Liposome preparation comprising a senescent cell derived inhibitor
PT   - use in the treatment of cancer, glaucoma, skin diseases and as an
PT   anti-ageing formula, etc.
XX
PS   Example 1; Page 154-155; 193pp; English.
XX
CC   A cDNA clone (AAT18792), designated SDI-1, codes for a senescent cell-
CC   derived inhibitor (AAR94932) that plays a crucial role in the
CC   expression of the senescent phenotype. The cDNA was identified in a
CC   library derived from quiescent normal human neonatal foreskin
CC   fibroblasts by transfecting the library into young, cycling cells
CC   and identifying clones that suppressed the initiation of DNA
CC   synthesis. Expression of SDI-1 increases 20-fold at cellular
CC   senescence. The cDNA, or the expressed protein, can be incorporated
CC   into a liposome and used to treat undesired cell proliferation, e.g.
CC   to treat cancer. Antisense sequences may be used to treat undesired
CC   cellular quiescence. Assays of cellular SDI-1 expression can be
CC   used to diagnose the presence and severity of p53-dependent cancers.
XX
SQ   Sequence 2106 BP; 404 A; 632 C; 575 G; 495 T; 0 other;

Alignment Scores:
Pred. No.:              198          Length:         2106
Score:                  38.00        Matches:        7
Percent Similarity:     100.00%      Conservative:   1
Best Local Similarity:  87.50%       Mismatches:     0
Query Match:            92.68%       Indels:         0
DB:                     17           Gaps:           0

US-09-726-470A-35 (1-8) x AAT18792 (1-2106)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    532 CACTCCAAACGCCGGCTGATCTTC 555

RESULT 8
AAT06940
ID   AAT06940 standard; DNA; 2106 BP.
XX
AC   AAT06940;
XX
DT   27-JUN-1996  (first entry)
XX
DE   Senescent cell derived inhibitor-1 coding sequence.
XX
KW   Senescent cell derived inhibitor-1; SCDI-1; mimetic; inhibitor; CDK;
KW   cyclin-dependent kinase; therapy; senescent cell; quiescent cell; tumour;
KW   progeria; Alzheimer's disease; astheria; cachexia; viral infection; CDK2;
KW   fungal infection; yeast infection; protozoan infection; fertility;
KW   helminthic infection; nematodal infection; parasitic infection; burn;
KW   wound healing; angiogenesis; endothelial cell proliferation; ageing;
KW   tissue degeneration; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             79..583
FT                   /*tag=  a
XX
PN   WO9531995-A1.
XX
PD   30-NOV-1995.
XX
PF   23-MAY-1995;    95WO-US06451.
XX
PR   24-OCT-1994;    94US-0327874.
PR   24-MAY-1994;    94US-0249371.
PR   30-JUN-1994;    94US-0268439.
PR   13-JUL-1994;    94US-0274535.
PR   26-AUG-1994;    94WO-US09700.
PR   03-OCT-1994;    94US-0321814.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
PA   (UYNC-) UNIV NORTH CAROLINA.
XX
PI   Kay BK,  Smith JR;
XX
DR   WPI; 1996-020353/02.
DR   P-PSDB; AAR86782.
XX
PT   DNA encoding mimetic(s) of the senescent cell derived inhibitor-1 -
PT   used for inhibition of DNA synthesis in active cells or suppressing
PT   such inhibition in senescent or quiescent cells
XX
PS   Disclosure; Fig 5; 63pp; English.
XX
CC   This sequence represents the coding sequence for senescent cell derived
CC   inhibitor-1 (SDCI-1) protein. Mimetic sequences (see AAR86772-R86778)
CC   that exhibit inhibition of SDI-1 activity can be created. The mimetcs
CC   are also capable of binding to a cyclin-dependent kinase (CDK),
CC   preferrably CDK2. The mimetic sequences can be used for diagnostic,
CC   therapeutic or experimental purposes, e.g. for inducing the inhibition of
CC   DNA synthesis in active cells, for suppressing such inhibition in
CC   senescent or quiescent cells. The therapeutic purposes include treating
CC   tumours, progeria, age related disorders (e.g. Alzheimer's disease,
CC   astheria and cachexia) and to treat viral, fungal, yeast, protozoan,
CC   helminthic, nematodal and other parasitic infections. The mimetics can
CC   also be used to increase fertility, or to induce fertility, to promote
CC   wound healing (angiogenesis, endothelial cell proliferation and recovery
CC   from burns), to create and study animal models for ageing, disease or
CC   tissue degeneration, to produce permanent cell lines for the treatment of
CC   cells ex vivo for re-implantation.
XX
SQ   Sequence 2106 BP; 401 A; 636 C; 574 G; 495 T; 0 other;

Alignment Scores:
Pred. No.:              198          Length:         2106
Score:                  38.00        Matches:        7
Percent Similarity:     100.00%      Conservative:   1
Best Local Similarity:  87.50%       Mismatches:     0
Query Match:            92.68%       Indels:         0
DB:                     17           Gaps:           0

US-09-726-470A-35 (1-8) x AAT06940 (1-2106)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    532 CACTCCAAACGCCGGCTGATCTTC 555

RESULT 9
AAQ90445
ID   AAQ90445 standard; cDNA; 2121 BP.
XX
AC   AAQ90445;
XX
DT   21-JAN-1996  (first entry)
XX
DE   Human WAF1 gene cDNA.
XX
KW   WAF1; p53; tummor suppressor gene; brain tumor; cancer therapy;
KW   cancer diagnosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             76..568
FT                   /*tag=  a
XX
PN   WO9513375-A1.
XX
PD   18-MAY-1995.
XX
PF   10-NOV-1994;    94WO-US12936.
XX
PR   10-NOV-1993;    93US-0149829.
```

```
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Kinzler KW,  Vogelstein B;
XX
DR   WPI; 1995-194094/25.
DR   P-PSDB; AAR73994.
XX
PT   Human gene WAF1, induced by wild-type p53 in human brain tumour
PT   cells  - also protein, antibodies and constructs useful for the
PT   treatment and diagnosis of human tumours
XX
PS   Claim 1; Page 37; 63pp; English.
XX
CC   This WAF1 gene is inducible by wild-type but not by mutant p53 in
CC   human brain tumor cells.  The sequence can be used for the generation
CC   of probes, especially for cancer diagnosis by comparing the DNA of
CC   normal and tumor cells to see if a mutant is present in the tumor
CC   cell.  The sequence can also be used to assess the susceptibility to
CC   cancers, by testing a tissue, e.g. blood, chorionic villi, amniotic
CC   fluid or a blastomere of a pre-implantation embryo, to determine
CC   whether DNA in the tissue contains a mutant WAF1 gene.  The DNA is
CC   no more than 90 kb in size and is a P1 clone.
XX
SQ   Sequence 2121 BP; 418 A; 628 C; 575 G; 500 T; 0 other;

Alignment Scores:
Pred. No.:              200           Length:        2121
Score:                  38.00         Matches:       7
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  87.50%        Mismatches:    0
Query Match:            92.68%        Indels:        0
DB:                     16            Gaps:          0

US-09-726-470A-35 (1-8) x AAQ90445 (1-2121)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db    529 CACTCCAAACGCCGGCTGATCTTC 552

RESULT 10
AAT61419
ID   AAT61419 standard; cDNA; 2121 BP.
XX
AC   AAT61419;
XX
DT   23-MAY-1997  (first entry)
XX
DE   Human WAF1 cDNA.
XX
KW   WAF1; wild-type p53 activated fragment 1; Cip1;
KW   CDK-interacting protein-1; SDI1; senescent cell-derived inhibitor;
KW   p21; cyclin-dependent kinase inhibitor protein; antisense;
KW   neuroblastoma; melanoma; epithelioma; fibroblastoma; carcinoma;
KW   leukaemia; myeloma; cancer; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             76..570
FT                   /*tag=  a
XX
PN   WO9703681-A1.
XX
PD   06-FEB-1997.
XX
PF   19-JUL-1996;   96WO-US11886.
XX
PR   20-JUL-1995;   95US-0001248.
XX
PA   (WORC-) WORCESTER FOUND BIOMEDICAL RES.
XX
PI   Poluha DK,  Poluha W,  Ross AH;
XX
DR   WPI; 1997-132367/12.
DR   P-PSDB; AAW13655.
XX
PT   Use of wild-type p53 activated fragment 1 inhibitors - for killing
PT   or inhibiting the growth of cells in which a WAF1-dependent pathway
PT   has been induced
XX
PS   Claim 4; Page 21-24; 32pp; English.
XX
CC   A cDNA clone (AAT61419) codes for wild-type p53 activated fragment 1
CC   (WAF1) (AAW13655), a cyclin-dependent kinase inhibitor which appears
CC   to be involved in the arrest of the cell cycle at a checkpoint in
CC   G1.  Methods of killing or inhibiting the growth of cells in which
CC   the WAF1 gene is being expressed comprise the administration of a
CC   WAF1 inhibitor, such as a WAF1-antisense oligonucleotide (see also
CC   AAT13844) or a vector which expresses the antisense oligonucleotide.
CC   The cells to be treated are pref. cancer cells in a human host,
CC   e.g. neuroblastoma, melanoma, epithelioma, fibroblastoma,
CC   carcinoma, leukaemia and myeloma cells.
XX
SQ   Sequence 2121 BP; 418 A; 628 C; 575 G; 500 T; 0 other;

Alignment Scores:
Pred. No.:              200           Length:        2121
Score:                  38.00         Matches:       7
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  87.50%        Mismatches:    0
Query Match:            92.68%        Indels:        0
DB:                     18            Gaps:          0

US-09-726-470A-35 (1-8) x AAT61419 (1-2121)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db    529 CACTCCAAACGCCGGCTGATCTTC 552

RESULT 11
AAX15105
ID   AAX15105 standard; cDNA; 2121 BP.
XX
AC   AAX15105;
XX
DT   14-APR-1999  (first entry)
XX
DE   cDNA encoding a protein designated p21-WAF1.
XX
KW   Cyclin-dependent kinase inhibitor; p21-WAF1; in vitro gene expression;
KW   transcriptional regulatory region; diagnosis; gastrointestinal cancer;
KW   ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             76..570
FT                   /*tag=  a
FT                   /product= p21-WAF1
XX
PN   US5871968-A.
XX
PD   16-FEB-1999.
XX
PF   05-FEB-1997;   97US-0795015.
XX
PR   18-DEC-1995;   95US-0574043.
PR   10-NOV-1993;   93US-0149829.
PR   05-FEB-1997;   97US-0795015.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   El-Deiry W,  Kinzler KW,  Vogelstein B;
XX
DR   WPI; 1999-166643/14.
```

```
DR   P-PSDB; AAW96746.
XX
PT   Expressing genes in cell that express p21WAF1  - for use in gene
PT   therapy and for the diagnosis of gastrointestinal cancers
XX
PS   Disclosure; Columns 19-22; 30pp; English.
XX
CC   The present sequence encodes a cyclin-dependent kinase inhibitor protein
CC   designated p21-WAF1. The specification describes a method for in
CC   vitro expression of a gene in a cell that expresses p21-WAF1. The method
CC   comprises administering to the cell a nucleic acid construct containing
CC   the p21WAF1 transcriptional regulatory region linked in cis configuration
CC   to the gene that is to be expressed. The method is used the diagnosis of
CC   gastrointestinal cancers.
XX
SQ   Sequence 2121 BP; 418 A; 628 C; 575 G; 500 T; 0 other;

Alignment Scores:
Pred. No.:                200          Length:         2121
Score:                    38.00        Matches:        7
Percent Similarity:       100.00%      Conservative:   1
Best Local Similarity:    87.50%       Mismatches:     0
Query Match:              92.68%       Indels:         0
DB:                       20           Gaps:           0

US-09-726-470A-35 (1-8) x AAX15105 (1-2121)

Qy       1 HisAlaLysArgArgLeuIlePhe 8
           |||:::|||||||||||||||||||
Db     529 CACTCCAAACGCCGGCTGATCTTC 552

RESULT 12
ABK84187
ID   ABK84187 standard; cDNA; 2121 BP.
XX
AC   ABK84187;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #758.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US30821.
XX
PR   03-OCT-2000; 2000US-237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression
PT   of genes associated with granulocyte activation, which serves as
PT   diagnostic markers that is useful for monitoring disease states and
PT   drug toxicity  -
XX
PS   Claim 1; SEQ ID No 758; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing
CC   the expression level to an expression level in an unactivated
CC   GC, where differential expression of Gs is indicative of GCA.
CC   Also included are modulating (M2) GA by contacting GC with an agent
CC   that alters the expression of at least one gene in Gs; (2) screening (M3)
CC   for an agent capable of modulating GCA or an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease using the
CC   gene expression profile; (3) detecting (M4) an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease, by detecting the
CC   level of expression in a sample of the tissue of gene(s) from Gs, where
CC   the level of expression of the gene is indicative of inflammation;
CC   (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC   an allergic response in a subject, exposure of a subject to a pathogen
CC   or sterile inflammatory disease, by contacting a tissue having
CC   inflammation with an agent that modulates the expression of gene(s)
CC   from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC   modulating GA; M3 is useful for screening an agent capable of modulating
CC   GCA preferably in an inflammation in a tissue; M4 is useful for
CC   detecting an inflammation (especially chronic) in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC   glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC   reperfusion injury, ARDS, adult respiratory distress syndrome,
CC   inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC   periodontal disease; also bacterial infection, viral infection,
CC   parasitic infection, protozoal infection, fungal infection and M5 is
CC   useful for treating one of the above conditions. The present
CC   sequence represents a gene differentially expressed in granulocytes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2121 BP; 418 A; 628 C; 575 G; 500 T; 0 other;

Alignment Scores:
Pred. No.:                200          Length:         2121
Score:                    38.00        Matches:        7
Percent Similarity:       100.00%      Conservative:   1
Best Local Similarity:    87.50%       Mismatches:     0
Query Match:              92.68%       Indels:         0
DB:                       24           Gaps:           0

US-09-726-470A-35 (1-8) x ABK84187 (1-2121)

Qy       1 HisAlaLysArgArgLeuIlePhe 8
           |||:::|||||||||||||||||||
Db     529 CACTCCAAACGCCGGCTGATCTTC 552

RESULT 13
AAI72397
ID   AAI72397 standard; cDNA; 2121 BP.
XX
AC   AAI72397;
XX
DT   02-MAY-2002  (first entry)
XX
DE   p21-Cip1 cDNA.
XX
KW   Cell cycle inhibitor; antisense; inner ear; sensory hair cell;
KW   support cell; auditory function; hearing disorder;
KW   sensory neuronal hearing loss; SNHL; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             76..570
FT                   /*tag=  a
FT                   /product= "p21-Cip1"
```

XX
PN WO200204605-A2.
XX
PD 17-JAN-2002.
XX
PF 10-JUL-2001; 2001WO-US21793.
XX
PR 11-JUL-2000; 2000US-0614099.
XX
PA (OTOG-) OTOGENE USA INC.
PA (OTOG-) OTOGENE AG.
XX
PI Kil J, Gu R, Grigeur C, Lowenheim H;
XX
DR WPI; 2002-171713/22.
DR P-PSDB; AAB47881.
XX
PT Stimulating the formation of inner ear sensory hair cells, useful for
PT treating hearing disorder involves damaging first inner ear sensory
PT hair cells and promoting the formation of new sensory hair cells from
PT inner ear support cells -
XX
PS Claim 18; Page 65-67; 77pp; English.
XX
CC The sequences given in AAI72395-401 encode cell cycle inhibitors.
CC These nucleic acids may be hybridised by antisense molecules in the
CC method of the invention. The method is for stimulating the formation
CC of an inner ear sensory hair cell from an inner ear support cell and
CC involves damaging a first inner ear sensory hair cell under conditions
CC that promote the formation of at least one inner ear sensory hair cell
CC that is in contact with the damaged first inner ear hair cell.
CC The method is useful for stimulating the formation of inner ear
CC cells e.g. sensory hair cells and support cells, for improving an
CC auditory function in an inner ear, in the treatment of hearing disorder
CC e.g. sensory neuronal hearing loss (SNHL), to identify genes and/or
CC proteins that are capable of stimulating the formation of inner ear
CC sensory hair cells and/or the formation of inner ear support cells
CC from sensory hair cells. The method damages and/or kills the inner
CC ear sensory cells, such as sensory hair cells and support cells, which
CC results in the increased stimulation in the formation of new, inner ear
CC hair cells, thus resulting in the improved curing of the auditory
CC function.
XX
SQ Sequence 2121 BP; 418 A; 628 C; 575 G; 500 T; 0 other;

Alignment Scores:

Pred. No.:	200	Length:	2121
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	92.68%	Indels:	0
DB:	24	Gaps:	0

US-09-726-470A-35 (1-8) x AAI72397 (1-2121)

```
Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    529 CACTCCAAACGCCGGCTGATCTTC 552
```

RESULT 14
AAS94878
ID AAS94878 standard; DNA; 2127 BP.
XX
AC AAS94878;
XX
DT 14-FEB-2002 (first entry)
XX
DE Human DNA sequence #133 expressed during foam cell differentiation.
XX
KW Human; foam cell differentiation; atherosclerosis; cerebral stroke;
KW cardiovascular disorder; coronary artery disease; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200177389-A2.
XX
PD 18-OCT-2001.
XX
PF 04-APR-2001; 2001WO-US11128.
XX
PR 05-APR-2000; 2000US-195106P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Shiffman D, Somogyi R, Lawn R, Seilhamer JJ, Porter GJ, Mikita T;
PI Tai J;
XX
DR WPI; 2002-010925/01.
XX
PT Composition useful for diagnosis of conditions, disorders or diseases
PT associated with atherosclerosis, comprises several polynucleotides that
PT are differentially expressed in foam cell development -
XX
PS Claim 1; Page 189; 315pp; English.
XX
CC The present invention relates to the isolation of human polynucleotide
CC sequences that are differentially expressed during foam cell
CC differentiation. The polynucleotide sequences of the invention or a
CC composition comprising these polynucleotides are useful as a high
CC throughput method for detecting altered expression of one or more
CC polynucleotides in a sample. The polynucleotides can be used in the
CC diagnosis of disorders associated with foam cell development such as
CC atherosclerosis, cerebral stroke, and cardiovascular disorders such as
CC coronary artery disease. The polynucleotide sequences can also be used
CC as PCR primers and probes. The polynucleotides of the invention are also
CC useful in gene therapy. AAS94746-AAS95021 represent the human
CC polynucleotide sequences of the invention which are differentially
CC expressed during foam cell differentiation.
XX
SQ Sequence 2127 BP; 409 A; 633 C; 583 G; 502 T; 0 other;

Alignment Scores:

Pred. No.:	200	Length:	2127
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	92.68%	Indels:	0
DB:	24	Gaps:	0

US-09-726-470A-35 (1-8) x AAS94878 (1-2127)

```
Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    548 CACTCCAAACGCCGGCTGATCTTC 571
```

RESULT 15
AAQ90051
ID AAQ90051 standard; cDNA; 2147 BP.
XX
AC AAQ90051;
XX
DT 27-OCT-1995 (first entry)
XX
DE Melanoma differentiation associated gene mda-6.
XX
KW Melanoma differentiation associated gene; mda; cancer;
KW HO-1 melanoma cell; p21; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 95..589
FT /*tag= a
XX
PN WO9511986-A.
XX

PD 04-MAY-1995.
XX
PF 24-OCT-1994; 94WO-US12160.
XX
PR 30-SEP-1994; 94US-0316537.
PR 27-OCT-1993; 93US-0143576.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Fisher PB, Jiang H;
XX
DR WPI; 1995-178878/23.
DR P-PSDB; AAR74226.
XX
PT New method for generation of subtracted cDNA libraries - used
PT partic. for identifying melanoma differentiation associated genes
PT for cancer study, diagnosis and therapy.
XX
PS Claim 56; Figure 23; 421pp; English.
XX
CC mda-6 is identical to WAF1, CIP1, SDI1 that encodes a Mr. 21,000
CC protein (p21). It represents a novel gene which is enhanced in
CC HO-1 cell by IFN-beta + MEZ. It represents a terminal
CC differentiation-regulated gene displaying increased expression in
CC all melanomas tested, in specific carcinomas, in normal cerebellum
CC cells and in GBM cells treated with IFN-beta + MEZ. The plasmid
CC mda-6 (ATCC 75585) is claimed.
XX
SQ Sequence 2147 BP; 430 A; 633 C; 583 G; 501 T; 0 other;

Alignment Scores:

Pred. No.:	202	Length:	2147
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	92.68%	Indels:	0
DB:	16	Gaps:	0

US-09-726-470A-35 (1-8) x AAQ90051 (1-2147)

```
Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db    548 CACTCCAAACGCCGGCTGATCTTC 571
```

Search completed: December 14, 2002, 16:00:08
Job time : 223 secs

GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2002, 15:51:10 ; Search time 1552 Seconds
(without alignments)
83.482 Million cell updates/sec

Title: US-09-726-470A-35
Perfect score: 41
Sequence: 1 HAKRRLIF 8

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09726470/runat_10122002_090717_4984/app_query.fasta_1.398
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09726470_@CGN_1_1_1716_@runat_10122002_090717_4984 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	38	92.7	131	10	AW843743	AW843743 CM4-CN004
	2	38	92.7	164	9	AA065009	AA065009 zm12b07.r
	3	38	92.7	164	9	AA310434	AA310434 EST18125
c	4	38	92.7	209	10	AW842793	AW842793 MR2-CN003
	5	38	92.7	225	14	BM753145	BM753145 K-EST0029
	6	38	92.7	250	9	AA375906	AA375906 EST88256
	7	38	92.7	250	9	AA376396	AA376396 EST88803
	8	38	92.7	257	14	R46847	R46847 yj54a03.r1
	9	38	92.7	259	10	BB372883	BB372883 BB372883
c	10	38	92.7	280	10	AW843746	AW843746 CM4-CN004
	11	38	92.7	303	9	AA376199	AA376199 EST88590
	12	38	92.7	334	14	D31116	D31116 HUML12552 H
	13	38	92.7	344	13	BI012565	BI012565 QV0-EN010
	14	38	92.7	348	14	BM752981	BM752981 K-EST0029
	15	38	92.7	373	13	BI035902	BI035902 IL5-NT022
	16	38	92.7	378	12	BF738945	BF738945 PM3-KT000
c	17	38	92.7	415	10	AW820448	AW820448 QV2-ST029
	18	38	92.7	445	13	BG981847	BG981847 IL3-CN010
	19	38	92.7	447	14	BM753949	BM753949 K-EST0031
	20	38	92.7	467	14	BM753774	BM753774 K-EST0030
c	21	38	92.7	479	9	AA029109	AA029109 zk10f07.s
	22	38	92.7	479	13	BG981844	BG981844 IL3-CN010
	23	38	92.7	482	14	BM748126	BM748126 K-EST0022
	24	38	92.7	504	10	BE206752	BE206752 ba02e12.y
	25	38	92.7	519	14	BM742288	BM742288 K-EST0015
	26	38	92.7	539	10	BE014412	BE014412 126115 MA
	27	38	92.7	541	9	AA029873	AA029873 zk10f07.r
	28	38	92.7	543	10	BE263622	BE263622 601192045
	29	38	92.7	547	10	AW239199	AW239199 xb37c04.y
	30	38	92.7	556	10	BE297240	BE297240 601177912
	31	38	92.7	572	10	AW247234	AW247234 2820687.5
	32	38	92.7	574	9	AA481712	AA481712 zv45g04.r
	33	38	92.7	575	10	AW250360	AW250360 2822083.5
	34	38	92.7	576	10	AW732606	AW732606 bb09a12.y
	35	38	92.7	578	10	AW249122	AW249122 2820943.5
	36	38	92.7	581	10	BE252329	BE252329 601114191
	37	38	92.7	582	13	BI117867	BI117867 602866893
c	38	38	92.7	585	14	BQ369014	BQ369014 PM3-GN051
c	39	38	92.7	585	14	BQ369197	BQ369197 PM3-GN051
	40	38	92.7	590	10	BE206983	BE206983 ba07c11.y
	41	38	92.7	591	10	AW952075	AW952075 EST364145
	42	38	92.7	604	14	BM753704	BM753704 K-EST0030
	43	38	92.7	605	14	BM783403	BM783403 K-EST0061
	44	38	92.7	606	10	BE255900	BE255900 601109857
	45	38	92.7	609	10	BE279288	BE279288 601157617

ALIGNMENTS

RESULT 1
AW843743
LOCUS AW843743 131 bp mRNA linear EST 18-MAY-2000
DEFINITION CM4-CN0043-120100-075-f04 CN0043 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW843743
VERSION AW843743.1 GI:7937726
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 131)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

```
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM4-CN0043-120
            100-075-f04&t3=2000-01-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 75
            High quality sequence stop: 131.
FEATURES             Location/Qualifiers
     source          1. .131
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CN0043"
                     /dev_stage="Adult"
                     /note="Organ: colon_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       27 a     41 c     35 g     28 t
ORIGIN

Alignment Scores:
Pred. No.:                26.2          Length:         131
Score:                    38.00         Matches:        7
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    87.50%        Mismatches:     0
Query Match:              92.68%        Indels:         0
DB:                       10            Gaps:           0

US-09-726-470A-35 (1-8) x AW843743 (1-131)

Qy       1 HisAlaLysArgArgLeuIlePhe 8
           |||:::||||||||||||||||||
Db      82 CACTCCAAACGCCGGCTGATCTTC 105

RESULT 2
AA065009
LOCUS       AA065009                 164 bp    mRNA    linear   EST 23-DEC-1997
DEFINITION  zm12b07.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
            IMAGE:525397 5' similar to SW:CDN1_HUMAN P38936 CYCLIN-DEPENDENT
            KINASE INHIBITOR 1 ;, mRNA sequence.
ACCESSION   AA065009
VERSION     AA065009.1  GI:1558625
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            WARNING: There is evidence that suggests that the 384-well parent
            plate of this clone contains both human and mouse derived clones.
            Thus, the origin of this clone is uncertain. This caution should be
            kept in mind should you use this clone.

            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1573   Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 122.
FEATURES             Location/Qualifiers
     source          1. .164
                     /organism="Homo sapiens"
                     /db_xref="GDB:3916826"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:525397"
                     /clone_lib="Stratagene pancreas (#937208)"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Pancreatic adenocarcinoma cell line. Average
                     insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT       39 a     50 c     41 g     31 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:                35.1          Length:         164
Score:                    38.00         Matches:        7
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    87.50%        Mismatches:     0
Query Match:              92.68%        Indels:         0
DB:                       9             Gaps:           0

US-09-726-470A-35 (1-8) x AA065009 (1-164)

Qy       1 HisAlaLysArgArgLeuIlePhe 8
           |||:::||||||||||||||||||
Db      41 CACTCCAAACGCCGGCTGATCTTC 64

RESULT 3
AA310434
LOCUS       AA310434                 164 bp    mRNA    linear   EST 19-APR-1997
DEFINITION  EST18125 Heart I Homo sapiens cDNA 5' end similar to melanoma
            differentiation associated mRNA, mda-6, mRNA sequence.
ACCESSION   AA310434
VERSION     AA310434.1  GI:1962762
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164)
  AUTHORS   Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     Rapid cDNA sequencing (expressed sequence tags) from a
            directionally cloned human infant brain cDNA library
  JOURNAL   Nat. Genet. 4, 373-380 (1993)
  MEDLINE   94004965
COMMENT     Other_ESTs: EST18124 THC175238
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
```

```
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .164
                      /organism="Homo sapiens"
                      /db_xref="ATCC (inhost):116880"
                      /db_xref="taxon:9606"
                      /clone_lib="Heart I"
                      /sex="male"
                      /dev_stage="adult, 25 yrs"
                      /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
BASE COUNT       37 a     48 c     46 g     32 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                35.1          Length:         164
Score:                    38.00         Matches:        7
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    87.50%        Mismatches:     0
Query Match:              92.68%        Indels:         0
DB:                       9             Gaps:           0

US-09-726-470A-35 (1-8) x AA310434 (1-164)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    132 CACTCCAAACGCCGGCTGATCTTC 155

RESULT 4
AW842793/c
LOCUS       AW842793                209 bp    mRNA    linear   EST 18-MAY-2000
DEFINITION  MR2-CN0037-210200-101-h05 CN0037 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW842793
VERSION     AW842793.1  GI:7936776
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 209)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR2-CN0037-210
            200-101-h05&t3=2000-02-21&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 29
            High quality sequence stop: 209.
FEATURES              Location/Qualifiers
     source           1. .209
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="CN0037"
                      /dev_stage="Adult"
                      /note="Organ: colon_normal; Vector: puc18; Site_1: SmaI;
                      Site_2: SmaI; A mini-library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716 -  Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
                      low stringency conditions."
BASE COUNT       58 a     41 c     66 g     44 t
ORIGIN

Alignment Scores:
Pred. No.:                48.3          Length:         209
Score:                    38.00         Matches:        7
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    87.50%        Mismatches:     0
Query Match:              92.68%        Indels:         0
DB:                       10            Gaps:           0

US-09-726-470A-35 (1-8) x AW842793 (1-209)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    204 CACTCCAAACGCCGGCTGATCTTC 181

RESULT 5
BM753145
LOCUS       BM753145                225 bp    mRNA    linear   EST 04-MAR-2002
DEFINITION  K-EST0029952 S7SNU719 Homo sapiens cDNA clone S7SNU719-25-E12 5',
            mRNA sequence.
ACCESSION   BM753145
VERSION     BM753145.1  GI:19082763
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 225)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 25  row: E  column: 12
            High quality sequence stop: 225.
FEATURES              Location/Qualifiers
     source           1. .225
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="S7SNU719-25-E12"
                      /clone_lib="S7SNU719"
                      /sex="M"
                      /tissue_type="Stomach"
                      /cell_type="Epithelial"
                      /cell_line="SNU-719"
                      /lab_host="Top10F'"
                      /note="Organ: Stomach; Vector: pCNS; Site_1: EcoRI;
                      Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
                      bacterial alkaline phosphatase (BAP) and then decapped
                      with tabacco acid pyrophosphatase (TAP). The decapped
                      intact mRNA was ligated with DNA-RNA linker including EcoR
                      I site by treatment of T4 RNA ligase and the first strand
                      cDNA was synthesized from oligo dT-selected mRNA by
                      priming with dT-tailed vector. The dT-tailed vector was
```

adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

BASE COUNT 49 a 68 c 44 g 64 t
ORIGIN

Alignment Scores:

Pred. No.:	53.2	Length:	225
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	92.68%	Indels:	0
DB:	14	Gaps:	0

US-09-726-470A-35 (1-8) x BM753145 (1-225)

```
Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db     38 CACTCCAAACGCCGGCTGATCTTC 61
```

RESULT 6
AA375906
LOCUS AA375906 250 bp mRNA linear EST 21-APR-1997
DEFINITION EST88256 HSC172 cells II Homo sapiens cDNA 5' end similar to melanoma differentiation associated mRNA, mda-6, mRNA sequence.
ACCESSION AA375906
VERSION AA375906.1 GI:2028224
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 250)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT Other_ESTs: THC175238
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .250
/organism="Homo sapiens"
/db_xref="ATCC (inhost):180348"
/db_xref="taxon:9606"
/clone_lib="HSC172 cells II"
/cell_type="fibroblast"
/cell_line="HSC172 (60PDL)"
/dev_stage="fetal"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI ; Site_2: XhoI"
BASE COUNT 57 a 72 c 55 g 61 t 5 others
ORIGIN

Alignment Scores:

Pred. No.:	61.1	Length:	250
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	92.68%	Indels:	0
DB:	9	Gaps:	0

US-09-726-470A-35 (1-8) x AA375906 (1-250)

```
Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db     65 CACTCCAAACGCCGGCTGATCTTC 88
```

RESULT 7
AA376396
LOCUS AA376396 250 bp mRNA linear EST 21-APR-1997
DEFINITION EST88803 HSC172 cells II Homo sapiens cDNA 5' end similar to melanoma differentiation associated mRNA, mda-6, mRNA sequence.
ACCESSION AA376396
VERSION AA376396.1 GI:2028714
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 250)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT Other_ESTs: THC175238
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers

```
     source          1. .250
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):180846"
                     /db_xref="taxon:9606"
                     /clone_lib="HSC172 cells II"
                     /cell_type="fibroblast"
                     /cell_line="HSC172 (60PDL)"
                     /dev_stage="fetal"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
                     ; Site_2: XhoI"
BASE COUNT       56 a     70 c     76 g     42 t      6 others
ORIGIN

Alignment Scores:
Pred. No.:              61.1          Length:         250
Score:                  38.00         Matches:        7
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  87.50%        Mismatches:     0
Query Match:            92.68%        Indels:         0
DB:                     9             Gaps:           0

US-09-726-470A-35 (1-8) x AA376396 (1-250)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db    186 CACTCCAAACGNCGGCTGATCTTC 209

RESULT 8
R46847
LOCUS       R46847                  257 bp    mRNA    linear   EST 10-MAY-1995
DEFINITION  yj54a03.r1 Soares breast 2NbHBst Homo sapiens cDNA clone
            IMAGE:152524 5' similar to SP:S39358 S39358 CYCLIN KINASE INHIBITOR
            - ;, mRNA sequence.
ACCESSION   R46847
VERSION     R46847.1  GI:806244
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 257)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1831
            Source: IMAGE Consortium, LLNL This clone is available royalty-free
            through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov)
            for further information.
            Insert Length: 1831   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 258.
FEATURES              Location/Qualifiers
     source           1. .257
                      /organism="Homo sapiens"
                      /db_xref="GDB:564741"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:152524"
                      /clone_lib="Soares breast 2NbHBst"
                      /sex="Female"
                      /dev_stage="adult"
                      /lab_host="DH10B (ampicillin resistant)"
                      /note="Organ: breast; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of a modified pT7T3 vector (Pharmacia).
                      Library went through one round of normalization to a Cot =
                      230. Library constructed by Bento Soares and M.Fatima
                      Bonaldo."
BASE COUNT       56 a     67 c     65 g     64 t      5 others
ORIGIN

Alignment Scores:
Pred. No.:              63.4          Length:         257
Score:                  38.00         Matches:        7
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  87.50%        Mismatches:     0
Query Match:            92.68%        Indels:         0
DB:                     14            Gaps:           0

US-09-726-470A-35 (1-8) x R46847 (1-257)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db    100 CACTCCAAACGCCGGCTGATCTTC 123

RESULT 9
BB372883
LOCUS       BB372883                259 bp    mRNA    linear   EST 13-JUL-2000
DEFINITION  BB372883 RIKEN full-length enriched, 16 days embryo head Mus
            musculus cDNA clone C130067I09 3' similar to U09507 Mus musculus
            p21 (Waf1) mRNA, mRNA sequence.
ACCESSION   BB372883
VERSION     BB372883.1  GI:9085377
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 259)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
```

19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .259
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C130067I09"
/clone_lib="RIKEN full-length enriched, 16 days embryo head"
/sex="mixed"
/tissue_type="head"
/dev_stage="16 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"

BASE COUNT 53 a 88 c 37 g 81 t
ORIGIN

Alignment Scores:

Pred. No.:	64	Length:	259
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	92.68%	Indels:	0
DB:	10	Gaps:	0

US-09-726-470A-35 (1-8) x BB372883 (1-259)

```
Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db     49 CACTCCAAGCGCAGATTGATCTTC 72
```

RESULT 10
AW843746/c
LOCUS AW843746 280 bp mRNA linear EST 18-MAY-2000
DEFINITION CM4-CN0043-120100-075-a12 CN0043 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW843746
VERSION AW843746.1 GI:7937729
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 280)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM4-CN0043-120100-075-a12&t3=2000-01-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 37
High quality sequence stop: 280.

FEATURES Location/Qualifiers
source 1. .280
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CN0043"
/dev_stage="Adult"
/note="Organ: colon_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 56 a 84 c 74 g 66 t
ORIGIN

Alignment Scores:

Pred. No.:	71	Length:	280
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	92.68%	Indels:	0
DB:	10	Gaps:	0

US-09-726-470A-35 (1-8) x AW843746 (1-280)

```
Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db     74 CACTCCAAACGCCGGCTGATCTTC 51
```

RESULT 11
AA376199
LOCUS AA376199 303 bp mRNA linear EST 21-APR-1997
DEFINITION EST88590 HSC172 cells II Homo sapiens cDNA 5' end similar to melanoma differentiation associated mRNA, mda-6, mRNA sequence.
ACCESSION AA376199
VERSION AA376199.1 GI:2028519
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 303)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
TITLE Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT Other_ESTs: THC175238
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .303
/organism="Homo sapiens"
/db_xref="ATCC (inhost):180651"
/db_xref="taxon:9606"
/clone_lib="HSC172 cells II"
/cell_type="fibroblast"
/cell_line="HSC172 (60PDL)"
/dev_stage="fetal"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI ; Site_2: XhoI"
BASE COUNT 75 a 88 c 63 g 77 t
ORIGIN

Alignment Scores:

Pred. No.:	78.7	Length:	303
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	92.68%	Indels:	0
DB:	9	Gaps:	0

US-09-726-470A-35 (1-8) x AA376199 (1-303)

```
Qy       1 HisAlaLysArgArgLeuIlePhe 8
           |||:::|||||||||||||||||||
Db      46 CACTCCAAACGCCGGCTGATCTTC 69
```

RESULT 12
D31116
LOCUS D31116 334 bp mRNA linear EST 08-FEB-1995
DEFINITION HUML12552 Human fetal lung Homo sapiens cDNA 5', mRNA sequence.
ACCESSION D31116
VERSION D31116.1 GI:643996
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 334)
AUTHORS Sudo,K., Chinen,K. and Nakamura,Y.
TITLE 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA library
JOURNAL Genomics 24, 276-279 (1995)
MEDLINE 95213017
COMMENT Contact: Yusuke Nakamura
Institute of Medical Science
University of Tokyo
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
Tel: 81-3-5449-5372
Fax: 81-3-5449-5433
Email: yusuke@ims.u-tokyo.ac.jp.
FEATURES Location/Qualifiers
source 1. .334
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human fetal lung"
/note="Organ: ovary; Vector: Bluescript SK; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Total ovary tissue, normal, caucasian. Average insert size: 0.8 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT 82 a 91 c 74 g 84 t 3 others
ORIGIN

Alignment Scores:

Pred. No.:	89.4	Length:	334
Score:	38.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	87.50%	Mismatches:	0
Query Match:	92.68%	Indels:	0
DB:	14	Gaps:	0

US-09-726-470A-35 (1-8) x D31116 (1-334)

```
Qy       1 HisAlaLysArgArgLeuIlePhe 8
           |||:::|||||||||||||||||||
Db      59 CACTCCAAACGCCGGCTGATCTTC 82
```

RESULT 13
BI012565
LOCUS BI012565 344 bp mRNA linear EST 13-JUN-2001
DEFINITION QV0-EN0102-090401-520-c09 EN0102 Homo sapiens cDNA, mRNA sequence.
ACCESSION BI012565
VERSION BI012565.1 GI:14416636
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 344)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV0&t2=QV0-EN0102-090401-520-c09&t3=2001-04-09&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 344.
FEATURES Location/Qualifiers
source 1. .344
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="EN0102"
/dev_stage="Adult"
/note="Organ: lung_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 79 a 97 c 88 g 80 t
ORIGIN

```
Alignment Scores:
Pred. No.:                93            Length:        344
Score:                    38.00         Matches:       7
Percent Similarity:       100.00%       Conservative:  1
Best Local Similarity:    87.50%        Mismatches:    0
Query Match:              92.68%        Indels:        0
DB:                       13            Gaps:          0

US-09-726-470A-35 (1-8) x BI012565 (1-344)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    159 CACTCCAAACGCCGGCTGATCTTC 182

RESULT 14
BM752981
LOCUS       BM752981                348 bp    mRNA    linear   EST 04-MAR-2002
DEFINITION  K-EST0029737 S7SNU719 Homo sapiens cDNA clone S7SNU719-27-C01 5',
            mRNA sequence.
ACCESSION   BM752981
VERSION     BM752981.1  GI:19082599
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 348)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 27  row: C  column: 01
            High quality sequence stop: 348.
FEATURES             Location/Qualifiers
     source          1. .348
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="S7SNU719-27-C01"
                     /clone_lib="S7SNU719"
                     /sex="M"
                     /tissue_type="Stomach"
                     /cell_type="Epithelial"
                     /cell_line="SNU-719"
                     /lab_host="Top10F'"
                     /note="Organ: Stomach; Vector: pCNS; Site_1: EcoRI;
                     Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
                     bacterial alkaline phosphatase (BAP) and then decapped
                     with tabacco acid pyrophosphatase (TAP). The decapped
                     intact mRNA was ligated with DNA-RNA linker including EcoR
                     I site by treatment of T4 RNA ligase and the first strand
                     cDNA was synthesized from oligo dT-selected mRNA by
                     priming with dT-tailed vector. The dT-tailed vector was
                     adjusted to have about 60nt. The cDNA vector was
                     circularized with E. coli DNA ligase after digestion of
                     EcoRI which site is also included in vector. An RNA strand
                     converted to a DNA strand by Okayama-Berg method. The
                     obtained cDNA vectors were used for transformation of
                     competent cells E. coli Top10F' by electroporation method.
                     The cDNA libraries constructed by this method are
                     full-length enriched cDNA library."
BASE COUNT       85 a     97 c     67 g     99 t
ORIGIN

Alignment Scores:
Pred. No.:                94.4          Length:        348
Score:                    38.00         Matches:       7
Percent Similarity:       100.00%       Conservative:  1
Best Local Similarity:    87.50%        Mismatches:    0
Query Match:              92.68%        Indels:        0
DB:                       14            Gaps:          0

US-09-726-470A-35 (1-8) x BM752981 (1-348)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db     37 CACTCCAAACGCCGGCTGATCTTC 60

RESULT 15
BI035902
LOCUS       BI035902                373 bp    mRNA    linear   EST 14-JUN-2001
DEFINITION  IL5-NT0228-020101-377-g02 NT0228 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI035902
VERSION     BI035902.1  GI:14442528
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL5&t2=IL5-NT0228-
            020101-377-g02&t3=2001-01-02&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 343.
FEATURES             Location/Qualifiers
     source          1. .373
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0228"
                     /dev_stage="Adult"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       95 a     102 c     84 g     92 t
ORIGIN

Alignment Scores:
Pred. No.:                103           Length:        373
Score:                    38.00         Matches:       7
Percent Similarity:       100.00%       Conservative:  1
Best Local Similarity:    87.50%        Mismatches:    0
Query Match:              92.68%        Indels:        0
DB:                       13            Gaps:          0
```

US-09-726-470A-35 (1-8) x BI035902 (1-373)

```
Qy       1 HisAlaLysArgArgLeuIlePhe 8
             |||:::||||||||||||||||||
Db      83 CACTCCAAACGCCGGCTGATCTTC 106
```

Search completed: December 14, 2002, 17:45:58
Job time : 1556 secs

GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2002, 15:50:15 ; Search time 1607 Seconds
(without alignments)
144.880 Million cell updates/sec

Title: US-09-726-470A-35
Perfect score: 41
Sequence: 1 HAKRRLIF 8

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09726470/runat_10122002_090717_4972/app_query.fasta_1.398
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09726470_@CGN_1_1_3637_@runat_10122002_090717_4972 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	41	100.0	136241	2	AC120223	AC120223 Rattus no
	2	38	92.7	331	6	AX098478	AX098478 Sequence
	3	38	92.7	495	6	AR000108	AR000108 Sequence
	4	38	92.7	592	4	D84650	D84650 Felis catus
	5	38	92.7	600	9	S67388	S67388 p21=cyclin
	6	38	92.7	626	9	HUMCIP1WAF	L47232 Homo sapien
	7	38	92.7	626	9	HUMCIP1WAG	L47233 Homo sapien
	8	38	92.7	1194	6	AR206141	AR206141 Sequence
	9	38	92.7	2098	9	HUMCDKI	L25610 Homo sapien
	10	38	92.7	2098	9	HUMSDI1A	L26165 Human DNA s
	11	38	92.7	2106	6	AR060688	AR060688 Sequence
	12	38	92.7	2106	6	AR206136	AR206136 Sequence
	13	38	92.7	2114	9	BC000275	BC000275 Homo sapi
	14	38	92.7	2120	9	HSU09579	U09579 Human melan
	15	38	92.7	2121	6	AR035955	AR035955 Sequence
	16	38	92.7	2121	6	AR038841	AR038841 Sequence
	17	38	92.7	2121	6	AX376625	AX376625 Sequence
	18	38	92.7	2121	9	HSU03106	U03106 Human wild-
	19	38	92.7	2127	6	AX281724	AX281724 Sequence
	20	38	92.7	2128	9	BC001935	BC001935 Homo sapi
	21	38	92.7	2180	9	BC013967	BC013967 Homo sapi
	22	38	92.7	2274	9	BC000312	BC000312 Homo sapi
	23	38	92.7	10907	9	AF497972	AF497972 Homo sapi
c	24	38	92.7	145735	2	AC109601	AC109601 Oryza sat
c	25	38	92.7	179510	2	AC127421	AC127421 Mus muscu
	26	38	92.7	180668	2	AC020857	AC020857 Mus muscu
	27	38	92.7	195364	9	HS431A14	Z85996 Human DNA s
	28	37	90.2	494	10	AB017817	AB017817 Mus muscu
	29	37	90.2	494	10	AB017818	AB017818 Mus muscu
	30	37	90.2	727	10	BC002043	BC002043 Mus muscu
	31	37	90.2	733	10	MMU24173	U24173 Mus musculu
	32	37	90.2	759	10	RNU24174	U24174 Rattus norv
	33	37	90.2	810	10	RATCIP1A	L41275 Rattus norv
	34	37	90.2	855	10	MMU09507	U09507 Mus musculu
	35	37	90.2	1145	10	AF457187S2	AF457188 Mus muscu
c	36	37	90.2	1831	8	CSP420678	AJ420678 Chroomona
	37	37	90.2	42064	3	CBRG45K16	AC084643 Caenorhab
c	38	37	90.2	51680	9	AP000261	AP000261 Homo sapi
	39	37	90.2	94516	8	AP004526	AP004526 Lotus jap
	40	37	90.2	100000	9	AP000035	AP000035 Homo sapi
c	41	37	90.2	100000	9	AP000100	AP000100 Homo sapi
c	42	37	90.2	100000	9	AP000176	AP000176 Homo sapi
	43	37	90.2	105200	2	AC130924	AC130924 Rattus no
c	44	37	90.2	110000	2	AC125619_0	AC125619 Homo sapi
	45	37	90.2	145891	2	AC067898	AC067898 Homo sapi

ALIGNMENTS

RESULT 1

```
AC120223/c
LOCUS       AC120223              136241 bp    DNA     linear   HTG 18-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-307M22, *** SEQUENCING IN PROGRESS
            ***, 50 unordered pieces.
ACCESSION   AC120223
VERSION     AC120223.2  GI:21746475
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 136241)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 136241)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 136241)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 14, 2002 this sequence version replaced gi:20452765.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GUML
                Center clone name: CH230-307M22
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 92457 bases at least Q40
                Consensus quality: 97846 bases at least Q30
                Consensus quality: 101277 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 50 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1237: contig of 1237 bp in length
            *     1238     1337: gap of unknown length
            *     1338     2477: contig of 1140 bp in length
            *     2478     2577: gap of unknown length
            *     2578     3773: contig of 1196 bp in length
            *     3774     3873: gap of unknown length
            *     3874     5333: contig of 1460 bp in length
            *     5334     5433: gap of unknown length
            *     5434     6516: contig of 1083 bp in length
            *     6517     6616: gap of unknown length
            *     6617     7797: contig of 1181 bp in length
            *     7798     7897: gap of unknown length
            *     7898     9520: contig of 1623 bp in length
            *     9521     9620: gap of unknown length
            *     9621    11451: contig of 1831 bp in length
            *    11452    11551: gap of unknown length
            *    11552    12773: contig of 1222 bp in length
            *    12774    12873: gap of unknown length
            *    12874    14506: contig of 1633 bp in length
            *    14507    14606: gap of unknown length
            *    14607    15953: contig of 1347 bp in length
            *    15954    16053: gap of unknown length
            *    16054    18349: contig of 2296 bp in length
            *    18350    18449: gap of unknown length
            *    18450    19976: contig of 1527 bp in length
            *    19977    20076: gap of unknown length
            *    20077    21148: contig of 1072 bp in length
            *    21149    21248: gap of unknown length
            *    21249    22929: contig of 1681 bp in length
            *    22930    23029: gap of unknown length
            *    23030    25641: contig of 2612 bp in length
            *    25642    25741: gap of unknown length
            *    25742    28054: contig of 2313 bp in length
            *    28055    28154: gap of unknown length
            *    28155    30379: contig of 2225 bp in length
            *    30380    30479: gap of unknown length
            *    30480    32748: contig of 2269 bp in length
            *    32749    32848: gap of unknown length
            *    32849    35106: contig of 2258 bp in length
            *    35107    35206: gap of unknown length
            *    35207    37048: contig of 1842 bp in length
            *    37049    37148: gap of unknown length
            *    37149    39266: contig of 2118 bp in length
            *    39267    39366: gap of unknown length
            *    39367    40799: contig of 1433 bp in length
            *    40800    40899: gap of unknown length
            *    40900    43019: contig of 2120 bp in length
            *    43020    43119: gap of unknown length
            *    43120    45158: contig of 2039 bp in length
            *    45159    45258: gap of unknown length
            *    45259    47295: contig of 2037 bp in length
            *    47296    47395: gap of unknown length
```

```
            *    47396    48834: contig of 1439 bp in length
            *    48835    48934: gap of unknown length
            *    48935    51435: contig of 2501 bp in length
            *    51436    51535: gap of unknown length
            *    51536    54235: contig of 2700 bp in length
            *    54236    54335: gap of unknown length
            *    54336    57329: contig of 2994 bp in length
            *    57330    57429: gap of unknown length
            *    57430    59496: contig of 2067 bp in length
            *    59497    59596: gap of unknown length
            *    59597    61948: contig of 2352 bp in length
            *    61949    62048: gap of unknown length
            *    62049    63956: contig of 1908 bp in length
            *    63957    64056: gap of unknown length
            *    64057    66236: contig of 2180 bp in length
            *    66237    66336: gap of unknown length
            *    66337    69688: contig of 3352 bp in length
            *    69689    69788: gap of unknown length
            *    69789    72181: contig of 2393 bp in length
            *    72182    72281: gap of unknown length
            *    72282    75455: contig of 3174 bp in length
            *    75456    75555: gap of unknown length
            *    75556    78895: contig of 3340 bp in length
            *    78896    78995: gap of unknown length
            *    78996    82341: contig of 3346 bp in length
            *    82342    82441: gap of unknown length
            *    82442    84796: contig of 2355 bp in length
            *    84797    84896: gap of unknown length
            *    84897    88837: contig of 3941 bp in length
            *    88838    88937: gap of unknown length
            *    88938    93012: contig of 4075 bp in length
            *    93013    93112: gap of unknown length
            *    93113    97714: contig of 4602 bp in length
            *    97715    97814: gap of unknown length
            *    97815   100110: contig of 2296 bp in length
            *   100111   100210: gap of unknown length
            *   100211   105379: contig of 5169 bp in length
            *   105380   105479: gap of unknown length
            *   105480   111300: contig of 5821 bp in length
            *   111301   111400: gap of unknown length
            *   111401   116430: contig of 5030 bp in length
            *   116431   116530: gap of unknown length
            *   116531   121578: contig of 5048 bp in length
            *   121579   121678: gap of unknown length
            *   121679   128057: contig of 6379 bp in length
            *   128058   128157: gap of unknown length
            *   128158   136241: contig of 8084 bp in length.
FEATURES             Location/Qualifiers
     source          1. .136241
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-307M22"
BASE COUNT    39106 a  27003 c  27430 g  37671 t   5031 others
ORIGIN

Alignment Scores:
Pred. No.:              1.05e+03        Length:         136241
Score:                  41.00           Matches:        8
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-726-470A-35 (1-8) x AC120223 (1-136241)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          ||||||||||||||||||||||||
Db  91289 CATGCCAAGAGAAGATTGATCTTC 91266

RESULT 2
AX098478
LOCUS       AX098478                331 bp    DNA     linear   PAT 02-APR-2001
DEFINITION  Sequence 15 from Patent WO0120029.
ACCESSION   AX098478
VERSION     AX098478.1  GI:13537764
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 331)
  AUTHORS   Tocque,B., Bracco,L. and Schweighoffer,F.
  TITLE     Genetic markers of toxicity, preparation and uses thereof
  JOURNAL   Patent: WO 0120029-A 15 22-MAR-2001;
            Exonhit Therapeutics S.A. (FR)
FEATURES             Location/Qualifiers
     source          1. .331
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     mRNA            25. .228
                     /note="Identity to human cyclin-dependant kinase inhibitor
                     (WAF-1) mRNA cds and 3'utr, nucleotides 524-727. GenBank
                     Acc: L25610."
BASE COUNT       81 a     98 c     63 g     84 t      5 others
ORIGIN

Alignment Scores:
Pred. No.:              5.06            Length:         331
Score:                  38.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  87.50%          Mismatches:     0
Query Match:            92.68%          Indels:         0
DB:                     6               Gaps:           0

US-09-726-470A-35 (1-8) x AX098478 (1-331)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db     33 CACTCCAAACGCCGGTTGATCTTC 56

RESULT 3
AR000108
LOCUS       AR000108                495 bp    DNA     linear   PAT 04-DEC-1998
DEFINITION  Sequence 1 from patent US 5736318.
ACCESSION   AR000108
VERSION     AR000108.1  GI:3962639
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Munger,K. and Jones,D.Leanne.
  TITLE     Method and kit for evaluating human papillomavirus transformed
            cells
  JOURNAL   Patent: US 5736318-A 1 07-APR-1998;
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="unknown"
BASE COUNT       96 a    150 c    165 g     84 t
ORIGIN

Alignment Scores:
Pred. No.:              8.1             Length:         495
Score:                  38.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  87.50%          Mismatches:     0
Query Match:            92.68%          Indels:         0
DB:                     6               Gaps:           0

US-09-726-470A-35 (1-8) x AR000108 (1-495)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db    454 CACTCCAAACGCCGGCTGATCTTC 477

RESULT 4
```

```
D84650
LOCUS       D84650                  592 bp    mRNA    linear   MAM 31-OCT-1997
DEFINITION  Felis catus mRNA for p21/WAF1, complete cds.
ACCESSION   D84650
VERSION     D84650.1  GI:2588788
KEYWORDS    p21/WAF1.
SOURCE      Felis catus Lymph node cDNA to mRNA, clone:pCFW.31,pCFW.AR,and
            pCFW.AS.
  ORGANISM  Felis catus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
REFERENCE   1  (sites)
  AUTHORS   Okuda,M., Minehata,K., Setoguchi,A., Cho,K.-W., Nakamura,N.,
            Nishigaki,K., Watari,T., Cevario,S., O'Brien,S.J., Tsujimoto,H. and
            Hasegawa,A.
  TITLE     Cloning and chromosome mapping of the feline genes p21WAF1 and
            p27Kip1
  JOURNAL   Gene 198 (1-2), 141-147 (1997)
  MEDLINE   98036042
REFERENCE   2  (bases 1 to 592)
  AUTHORS   Okuda,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-1996) Masaru Okuda, The University of Tokyo,
            Veterinary Internal Medicine; Yayoi 1-1-1, Bunkyo-ku, Tokyo 113,
            Japan (E-mail:aa47111@hongo.ecc.u-tokyo.ac.jp,
            Tel:03-3812-2111(ex.5403), Fax:03-5800-6866)
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Felis catus"
                     /db_xref="taxon:9685"
                     /clone="pCFW.31,pCFW.AR,and pCFW.AS"
                     /tissue_type="Lymph node"
     CDS             74. .568
                     /codon_start=1
                     /product="p21/WAF1"
                     /protein_id="BAA23168.1"
                     /db_xref="GI:2588789"
                     /translation="MSEPSRDAHQIPHGSKACRRLFGPVDSEQLRRDCDALMAGCVQE
                     ARERWNFDFVTETPLEGDFAWERVWGLGLPKLYLPAGPRGGRDDLGGGKRPSTSSTLL
                     PGTAREDHLDLSLSCTLMPHSPERPEASPGAPGTSQGRKRRQTSMTDFYHSKRRLIFS
                     KRKP"
BASE COUNT      106 a    201 c    192 g     93 t
ORIGIN

Alignment Scores:
Pred. No.:              10              Length:         592
Score:                  38.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  87.50%          Mismatches:     0
Query Match:            92.68%          Indels:         0
DB:                     4               Gaps:           0

US-09-726-470A-35 (1-8) x D84650 (1-592)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db    527 CACTCCAAACGCCGGCTGATCTTC 550

RESULT 5
S67388
LOCUS       S67388                  600 bp    mRNA    linear   PRI 05-FEB-1994
DEFINITION  p21=cyclin kinase inhibitor [human, lung, WI38, mRNA, 600 nt].
ACCESSION   S67388
VERSION     S67388.1  GI:453134
KEYWORDS    .
SOURCE      Homo sapiens WI38 lung.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Xiong,Y., Hannon,G.J., Zhang,H., Casso,D., Kobayashi,R. and
            Beach,D.
  TITLE     p21 is a universal inhibitor of cyclin kinases
  JOURNAL   Nature 366 (6456), 701-704 (1993)
  MEDLINE   94081955
   PUBMED   8259214
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 140978] from the original journal article.
            This sequence comes from Fig. 1.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1. .600
                     /gene="p21"
     CDS             76. .570
                     /gene="p21"
                     /note="cyclin kinase inhibitor; Method: conceptual
                     translation with partial peptide sequencing; This sequence
                     comes from Fig. 1"
                     /codon_start=1
                     /product="p21"
                     /protein_id="AAB29246.1"
                     /db_xref="GI:453135"
                     /translation="MSEPAGDVRQNPCGSKACRRLFGPVDSEQLSRDCDALMAGCIQE
                     ARERWNFDFVTETPLEGDFAWERVRGLGLPKLYLPTGPRRGRDELGGGRRPGTSPALL
                     QGTAEEDHVDLSLSCTLVPRSGEQAEGSPGGPGDSQGRKRRQTSMTDFYHSKRRLIFS
                     KRKP"
BASE COUNT      116 a    183 c    203 g     98 t
ORIGIN

Alignment Scores:
Pred. No.:              10.2            Length:         600
Score:                  38.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  87.50%          Mismatches:     0
Query Match:            92.68%          Indels:         0
DB:                     9               Gaps:           0

US-09-726-470A-35 (1-8) x S67388 (1-600)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db    529 CACTCCAAACGCCGGCTGATCTTC 552

RESULT 6
HUMCIP1WAF
LOCUS       HUMCIP1WAF              626 bp    mRNA    linear   PRI 13-SEP-1995
DEFINITION  Homo sapiens cyclin-dependent kinase (CIP1/WAF1) mRNA, with a
            mutation (Ser 31 to Arg) predisposing to cancer, 3' end.
ACCESSION   L47232
VERSION     L47232.1  GI:984723
KEYWORDS    cyclin-dependent kinase; mutation.
SOURCE      Homo sapiens tumor cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Mousses,S., Ozcelik,H., Lee,P.D., Malkin,D., Bull,S.B. and
            Andrulis,I.L.
  TITLE     Two variants of the CIP1/WAF1 gene occur together and are
            associated with human cancer
  JOURNAL   Hum. Mol. Genet. 4 (6), 1089-1092 (1995)
  MEDLINE   95384154
   PUBMED   7655464
FEATURES             Location/Qualifiers
     source          1. .626
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.2"
                     /tissue_type="tumor"
     gene            1. .626
                     /gene="CIP1/WAF1"
     CDS             <1. .548
                     /gene="CIP1/WAF1"
                     /codon_start=3
```

```
                     /product="cyclin-dependent kinase"
                     /protein_id="AAB59559.1"
                     /db_xref="GI:984724"
                     /translation="RSWARIRRGTEALRGGAMSEPAGDVRQNPCGSKACRRLFGPVDS
                     EQLRRDCDALMAGCIQEARERWNFDFVTETPLEGDFAWERVRGLGLPKLYLPTGPRRG
                     RDELGGGRRPGTSPALLQGTAEEDHVDLSLSCTLVPRSGEQAEGSPGGPGDSQGRKRR
                     QTSMTDFYHSKRRLIFSKRKP"
     primer_bind     complement(1. .20)
     variation       146
                     /gene="CIP1/WAF1"
                     /note="c (Ser) in wt/a (Arg) in mutant"
                     /replace="c"
     primer_bind     607. .626
                     /gene="CIP1/WAF1"
BASE COUNT      121 a    194 c    208 g    103 t
ORIGIN

Alignment Scores:
Pred. No.:              10.7          Length:         626
Score:                  38.00         Matches:        7
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  87.50%        Mismatches:     0
Query Match:            92.68%        Indels:         0
DB:                     9             Gaps:           0

US-09-726-470A-35 (1-8) x HUMCIP1WAF (1-626)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db    507 CACTCCAAACGCCGGCTGATCTTC 530

RESULT 7
HUMCIP1WAG
LOCUS       HUMCIP1WAG              626 bp    mRNA    linear   PRI 15-SEP-1995
DEFINITION  Homo sapiens cyclin-dependent kinase (CIP1/WAF1) mRNA, 3' end, with
            a cancer predisposing mutation in the 3' UTR.
ACCESSION   L47233
VERSION     L47233.1  GI:986878
KEYWORDS    cyclin-dependent kinase; mutation.
SOURCE      Homo sapiens tumor cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Mousses,S., Ozcelik,H., Lee,P.D., Malkin,D., Bull,S.B. and
            Andrulis,I.L.
  TITLE     Two variants of the CIP1/WAF1 gene occur together and are
            associated with human cancer
  JOURNAL   Hum. Mol. Genet. 4 (6), 1089-1092 (1995)
  MEDLINE   95384154
   PUBMED   7655464
FEATURES             Location/Qualifiers
     source          1. .626
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="6p21.2"
                     /tissue_type="tumor"
     gene            1. .626
                     /gene="CIP1/WAF1"
     CDS             <1. .548
                     /gene="CIP1/WAF1"
                     /codon_start=3
                     /product="cyclin-dependent kinase"
                     /protein_id="AAB59560.1"
                     /db_xref="GI:986879"
                     /translation="RSWARIRRGTEALRGGAMSEPAGDVRQNPCGSKACRRLFGPVDS
                     EQLSRDCDALMAGCIQEARERWNFDFVTETPLEGDFAWERVRGLGLPKLYLPTGPRRG
                     RDELGGGRRPGTSPALLQGTAEEDHVDLSLSCTLVPRSGEQAEGSPGGPGDSQGRKRR
                     QTSMTDFYHSKRRLIFSKRKP"
     primer_bind     complement(1. .20)
     variation       568
                     /gene="CIP1/WAF1"
                     /note="c in wt/t in mutant"
                     /replace="c"
     primer_bind     607. .626
                     /gene="CIP1/WAF1"
BASE COUNT      120 a    194 c    208 g    104 t
ORIGIN

Alignment Scores:
Pred. No.:              10.7          Length:         626
Score:                  38.00         Matches:        7
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  87.50%        Mismatches:     0
Query Match:            92.68%        Indels:         0
DB:                     9             Gaps:           0

US-09-726-470A-35 (1-8) x HUMCIP1WAG (1-626)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db    507 CACTCCAAACGCCGGCTGATCTTC 530

RESULT 8
AR206141
LOCUS       AR206141                1194 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 10 from patent US 6372249.
ACCESSION   AR206141
VERSION     AR206141.1  GI:21504655
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1194)
  AUTHORS   Smith,J.R., Drutz,D.J., Wilson,D.R. and Zumstein,L.A.
  TITLE     Senscent cell-derived inhibitors of DNA synthesis
  JOURNAL   Patent: US 6372249-A 10 16-APR-2002;
FEATURES             Location/Qualifiers
     source          1. .1194
                     /organism="unknown"
BASE COUNT      298 a    278 c    321 g    297 t
ORIGIN

Alignment Scores:
Pred. No.:              22.7          Length:         1194
Score:                  38.00         Matches:        7
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  87.50%        Mismatches:     0
Query Match:            92.68%        Indels:         0
DB:                     6             Gaps:           0

US-09-726-470A-35 (1-8) x AR206141 (1-1194)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db   1153 CACTCCAAACGCCGGCTGATCTTC 1176

RESULT 9
HUMCDKI
LOCUS       HUMCDKI                 2098 bp    mRNA    linear   PRI 25-JAN-1994
DEFINITION  Homo sapiens cyclin-dependent kinase inhibitor mRNA, complete cds.
ACCESSION   L25610
VERSION     L25610.1  GI:425142
KEYWORDS    cyclin-dependent kinase inhibitor.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2098)
  AUTHORS   Harper,J.W., Adami,G.R., Wei,N., Keyomarsi,K. and Elledge,S.J.
  TITLE     The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1
            cyclin-dependent kinases
  JOURNAL   Cell 75 (4), 805-816 (1993)
  MEDLINE   94061996
   PUBMED   8242751
FEATURES             Location/Qualifiers
```

```
     source          1. .2098
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="B-lymphocyte (EBV transformed)"
     5'UTR           1. .78
     CDS             79. .573
                     /codon_start=1
                     /product="cyclin-dependent kinase inhibitor"
                     /protein_id="AAA16109.1"
                     /db_xref="GI:425143"
                     /translation="MSEPAGDVRQNPCGSKACRRLFGPVDSEQLSRDCDALMAGCIQE
                     ARERWNFDFVTETPLEGDFAWERVRGLGLPKLYLPTGPRRGRDELGGGRRPGTSPALL
                     QGTAEEDHVDLSLSCTLVPRSGEQAEGSPGGPGDSQGRKRRQTSMTDFYHSKRRLIFS
                     KRKP"
     3'UTR           572. .2098
     polyA_site      2098
BASE COUNT      396 a    632 c    575 g    495 t
ORIGIN

Alignment Scores:
Pred. No.:              44              Length:         2098
Score:                  38.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  87.50%          Mismatches:     0
Query Match:            92.68%          Indels:         0
DB:                     9               Gaps:           0

US-09-726-470A-35 (1-8) x HUMCDKI (1-2098)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    532 CACTCCAAACGCCGGCTGATCTTC 555

RESULT 10
HUMSDI1A
LOCUS       HUMSDI1A                2098 bp    mRNA    linear   PRI 18-JUL-1994
DEFINITION  Human DNA synthesis inhibitor mRNA, complete cds.
ACCESSION   L26165
VERSION     L26165.1  GI:418017
KEYWORDS    .
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2098)
  AUTHORS   Noda,A., Ning,Y., Venable,S.F., Pereira-Smith,O.M. and Smith,J.R.
  TITLE     Cloning of senescent cell-derived inhibitors of DNA synthesis using
            an expression screen
  JOURNAL   Exp. Cell Res. 211 (1), 90-98 (1994)
  MEDLINE   94170884
   PUBMED   8125163
FEATURES             Location/Qualifiers
     source          1. .2098
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="fibroblast"
     CDS             79. .573
                     /note="putative DNA synthesis inhibitor"
                     /codon_start=1
                     /protein_id="AAA19811.1"
                     /db_xref="GI:433742"
                     /translation="MSEPAGDVRQNPCGSKACRRLFGPVDSEQLSRDCDALMAGCIQE
                     ARERWNFDFVTETPLEGDFAWERVRGLGLPKLYLPTGPRRGRDELGGGRRPGTSPALL
                     QGTAEEDHVDLSLSCTLVPRSGEQAEGSPGGPGDSQGRKRRQTSMTDFYHSKRRLIFS
                     KRKP"
BASE COUNT      396 a    632 c    575 g    495 t
ORIGIN

Alignment Scores:
Pred. No.:              44              Length:         2098
Score:                  38.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  87.50%          Mismatches:     0
Query Match:            92.68%          Indels:         0
DB:                     9               Gaps:           0

US-09-726-470A-35 (1-8) x HUMSDI1A (1-2098)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    532 CACTCCAAACGCCGGCTGATCTTC 555

RESULT 11
AR060688
LOCUS       AR060688                2106 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1 from patent US 5840845.
ACCESSION   AR060688
VERSION     AR060688.1  GI:5987138
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2106)
  AUTHORS   Smith,J.R. and Noda,A.
  TITLE     Senescent cell derived inhibitors of DNA synthesis
  JOURNAL   Patent: US 5840845-A 1 24-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .2106
                     /organism="unknown"
BASE COUNT      404 a    637 c    570 g    495 t
ORIGIN

Alignment Scores:
Pred. No.:              44.2            Length:         2106
Score:                  38.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  87.50%          Mismatches:     0
Query Match:            92.68%          Indels:         0
DB:                     6               Gaps:           0

US-09-726-470A-35 (1-8) x AR060688 (1-2106)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    532 CACTCCAAACGCCGGCTGATCTTC 555

RESULT 12
AR206136
LOCUS       AR206136                2106 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6372249.
ACCESSION   AR206136
VERSION     AR206136.1  GI:21504649
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2106)
  AUTHORS   Smith,J.R., Drutz,D.J., Wilson,D.R. and Zumstein,L.A.
  TITLE     Senscent cell-derived inhibitors of DNA synthesis
  JOURNAL   Patent: US 6372249-A 1 16-APR-2002;
FEATURES             Location/Qualifiers
     source          1. .2106
                     /organism="unknown"
BASE COUNT      404 a    632 c    575 g    495 t
ORIGIN

Alignment Scores:
Pred. No.:              44.2            Length:         2106
Score:                  38.00           Matches:        7
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  87.50%          Mismatches:     0
Query Match:            92.68%          Indels:         0
DB:                     6               Gaps:           0

US-09-726-470A-35 (1-8) x AR206136 (1-2106)
```

```
Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db    532 CACTCCAAACGCCGGCTGATCTTC 555

RESULT 13
BC000275
LOCUS       BC000275               2114 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, cyclin-dependent kinase inhibitor 1A (p21, Cip1),
            clone MGC:3175 IMAGE:3355833, mRNA, complete cds.
ACCESSION   BC000275
VERSION     BC000275.1  GI:12653024
KEYWORDS    MGC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2114)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 6 Row: a Column: 4
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1. .2114
                     /organism="Homo sapiens"
                     /db_xref="LocusID:1026"
                     /db_xref="taxon:9606"
                     /clone="MGC:3175 IMAGE:3355833"
                     /tissue_type="Eye, retinoblastoma"
                     /clone_lib="NIH_MGC_16"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             66. .560
                     /codon_start=1
                     /product="cyclin-dependent kinase inhibitor 1A (p21,
                     Cip1)"
                     /protein_id="AAH00275.1"
                     /db_xref="GI:12653025"
                     /translation="MSEPAGDVRQNPCGSKACRRLFGPVDSEQLSRDCDALMAGCIQE
                     ARERWNFDFVTETPLEGDFAWERVRGLGLPKLYLPTGPRRGRDELGGGRRPGTSPALL
                     QGTAEEDHVDLSLSCTLVPRSGEQAEGSPGGPGDSQGRKRRQTSMTDFYHSKRRLIFS
                     KRKP"
BASE COUNT      418 a    627 c    571 g    498 t
ORIGIN

Alignment Scores:
Pred. No.:              44.4          Length:        2114
Score:                  38.00         Matches:       7
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  87.50%        Mismatches:    0
Query Match:            92.68%        Indels:        0
DB:                     9             Gaps:          0

US-09-726-470A-35 (1-8) x BC000275 (1-2114)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::||||||||||||||||||
Db    519 CACTCCAAACGCCGGCTGATCTTC 542

RESULT 14
HSU09579
LOCUS       HSU09579               2120 bp    mRNA    linear   PRI 26-JAN-1996
DEFINITION  Human melanoma differentiation associated (mda-6) mRNA, complete
            cds.
ACCESSION   U09579
VERSION     U09579.1  GI:495286
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2120)
  AUTHORS   Jiang,H. and Fisher,P.B.
  TITLE     Use of a sensitive and efficient subtraction hybridization protocol
            for the identification of genes differentially regulated during the
            induction of differentiation in human melanoma cells
  JOURNAL   Mol. Cell. Differ. 1, 285-299 (1993)
REFERENCE   2  (bases 1 to 2120)
  AUTHORS   Jiang,H., Lin,J., Su,Z.Z., Herlyn,M., Kerbel,R.S., Weissman,B.E.,
            Welch,D.R. and Fisher,P.B.
  TITLE     The melanoma differentiation-associated gene mda-6, which encodes
            the cyclin-dependent kinase inhibitor p21, is differentially
            expressed during growth, differentiation and progression in human
            melanoma cells
  JOURNAL   Oncogene 10 (9), 1855-1864 (1995)
  MEDLINE   95273102
   PUBMED   7753561
REFERENCE   3  (bases 1 to 2120)
  AUTHORS   Fisher,P.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-1994) Paul B. Fisher, Department of Pathology and
            Urology, Columbia University/College of Physicians and Surgeons,
            630 West 168th Street, New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1. .2120
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="p49c13, p49r3, p49r8, p49r23"
                     /cell_line="HO-1"
                     /cell_type="melanoma"
                     /clone_lib="cDNA library for differentiated HO-1 cells,
                     RACE library"
     gene            1. .2120
                     /gene="mda-6"
     CDS             95. .589
                     /gene="mda-6"
                     /note="alternate gene name=WAF1"
                     /codon_start=1
                     /evidence=experimental
                     /protein_id="AAA85641.1"
                     /db_xref="GI:495287"
                     /translation="MSEPAGDVRQNPCGSKACRRLFGPVDSEQLSRDCDALMAGCIQE
                     ARERWNFDFVTETPLEGDFAWERVRGLGLPKLYLPTGPRRGRDELGGGRRPGTSPALL
                     QGTAEEDHVDLSLSCTLVPRSGEQAEGSPGGPGDSQGRKRRQTSMTDFYHSKRRLIFS
                     KRKP"
     polyA_site      2120
                     /gene="mda-6"
                     /note="27 A residues"
BASE COUNT      403 a    633 c    583 g    501 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:              44.6          Length:        2120
Score:                  38.00         Matches:       7
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  87.50%        Mismatches:    0
Query Match:            92.68%        Indels:        0
DB:                     9             Gaps:          0

US-09-726-470A-35 (1-8) x HSU09579 (1-2120)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    548 CACTCCAAACGCCGGCTGATCTTC 571

RESULT 15
AR035955
LOCUS       AR035955               2121 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1 from patent US 5871968.
ACCESSION   AR035955
VERSION     AR035955.1  GI:5952623
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2121)
  AUTHORS   Kinzler,K.W., El-Deiry,W. and Vogelstein,B.
  TITLE     P21.sup.WAF1 derivatives and diagnostic methods
  JOURNAL   Patent: US 5871968-A 1 16-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .2121
                     /organism="unknown"
BASE COUNT      418 a    628 c    575 g    500 t
ORIGIN

Alignment Scores:
Pred. No.:              44.6          Length:        2121
Score:                  38.00         Matches:       7
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  87.50%        Mismatches:    0
Query Match:            92.68%        Indels:        0
DB:                     6             Gaps:          0

US-09-726-470A-35 (1-8) x AR035955 (1-2121)

Qy      1 HisAlaLysArgArgLeuIlePhe 8
          |||:::|||||||||||||||||||
Db    529 CACTCCAAACGCCGGCTGATCTTC 552


Search completed: December 14, 2002, 16:54:06
Job time : 1619 secs
```

GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2002, 15:51:10 ; Search time 1552 Seconds
(without alignments)
83.482 Million cell updates/sec

Title: US-09-726-470A-2
Perfect score: 20
Sequence: 1 XXXRXLXF 8

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool/US09726470/runat_10122002_090717_4984/app_query.fasta_1.398
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09726470_@CGN_1_1_1716_@runat_10122002_090717_4984 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	15	75.0	25	17	TA111D05P	AL461854 T. brucei
	2	15	75.0	30	17	AZ344289	AZ344289 1M0078P11
c	3	15	75.0	32	10	AV838306	AV838306 AV838306
c	4	15	75.0	33	17	BH792463	BH792463 SALK_0642
c	5	15	75.0	33	17	BH792464	BH792464 SALK_0642
	6	15	75.0	42	17	AQ024967	AQ024967 EP(2)0827
c	7	15	75.0	43	9	AA960099	AA960099 ub54b07.s
c	8	15	75.0	43	9	AI316449	AI316449 uj60d11.y
	9	15	75.0	44	17	AQ025720	AQ025720 l(2)k0320
	10	15	75.0	45	17	AQ025751	AQ025751 l(2)k0491
	11	15	75.0	47	17	AQ025803	AQ025803 l(2)k0661
c	12	15	75.0	48	17	AZ499421	AZ499421 1M0337I12
	13	15	75.0	50	9	AU105436	AU105436 AU105436
	14	15	75.0	50	9	AU105781	AU105781 AU105781
	15	15	75.0	50	9	AU105834	AU105834 AU105834
	16	15	75.0	50	9	AU106884	AU106884 AU106884
c	17	15	75.0	52	12	BF107513	BF107513 601823849
	18	15	75.0	53	14	BQ077176	BQ077176 fz13e08.y
	19	15	75.0	53	14	BQ077552	BQ077552 fz17h02.y
	20	15	75.0	53	14	BQ077596	BQ077596 fz18d04.y
	21	15	75.0	53	17	AL753521	AL753521 Arabidops
	22	15	75.0	54	17	AL755781	AL755781 Arabidops
	23	15	75.0	55	17	AZ825378	AZ825378 2M0100K03
c	24	15	75.0	56	10	AV840145	AV840145 AV840145
	25	15	75.0	56	14	BQ092961	BQ092961 fy92b12.y
	26	15	75.0	57	9	AA647971	AA647971 vq81f06.s
	27	15	75.0	57	9	AA681866	AA681866 vr44b01.s
	28	15	75.0	57	13	BI320187	BI320187 ic47g05.y
	29	15	75.0	57	17	TA93B08P	AL458792 T. brucei
c	30	15	75.0	58	9	AI128536	AI128536 qc68b04.x
c	31	15	75.0	58	12	BF457697	BF457697 UI-M-BZ1-
	32	15	75.0	59	10	AW546515	AW546515 L0009C05-
	33	15	75.0	59	17	AL751884	AL751884 Arabidops
	34	15	75.0	60	13	BI493320	BI493320 df99g12.y
c	35	15	75.0	60	17	AL768140	AL768140 Arabidops
c	36	15	75.0	61	9	AA796444	AA796444 vs95f03.r
	37	15	75.0	61	9	AI092739	AI092739 qa35e10.x
c	38	15	75.0	61	9	AI971846	AI971846 wv29b07.x
c	39	15	75.0	63	17	AZ922491	AZ922491 MRCot2E01
c	40	15	75.0	63	17	AZ922492	AZ922492 MRCot2B05
c	41	15	75.0	63	17	TA118D01P	AL463713 T. brucei
c	42	15	75.0	64	9	AA912565	AA912565 om52e12.s
c	43	15	75.0	64	14	F32526	F32526 HSPD25336 H
	44	15	75.0	64	17	AZ804082	AZ804082 2M0064J14
	45	15	75.0	64	17	FR0016325	AL007524 F.rubripe

ALIGNMENTS

RESULT 1
TA111D05P
LOCUS TA111D05P 25 bp DNA linear GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 111d05, forward sequence, genomic survey sequence.
ACCESSION AL461854
VERSION AL461854.1 GI:11832216
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
REFERENCE 1 (bases 1 to 25)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,

Melville,S.E., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nhl@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES Location/Qualifiers
source 1. .25
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="111d05"

BASE COUNT 6 a 8 c 2 g 9 t
ORIGIN

Alignment Scores:

Pred. No.:	5.77e+03	Length:	25
Score:	15.00	Matches:	3
Percent Similarity:	60.00%	Conservative:	0
Best Local Similarity:	60.00%	Mismatches:	2
Query Match:	75.00%	Indels:	0
DB:	17	Gaps:	0

US-09-726-470A-2 (1-8) x TA111D05P (1-25)

```
Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     10 CGTACCCTAACTTTT 24
```

RESULT 2
AZ344289
LOCUS AZ344289 30 bp DNA linear GSS 29-SEP-2000
DEFINITION 1M0078P11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0078P11 F, DNA sequence.
ACCESSION AZ344289
VERSION AZ344289.1 GI:10423377
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 30)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0078 row: P column: 11
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 30.

FEATURES Location/Qualifiers
source 1. .30
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0078P11"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 7 a 9 c 4 g 10 t
ORIGIN

Alignment Scores:

Pred. No.:	7.29e+03	Length:	30
Score:	15.00	Matches:	3
Percent Similarity:	60.00%	Conservative:	0
Best Local Similarity:	60.00%	Mismatches:	2
Query Match:	75.00%	Indels:	0
DB:	17	Gaps:	0

US-09-726-470A-2 (1-8) x AZ344289 (1-30)

```
Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     15 CGCACTCTTACTTTT 29
```

RESULT 3
AV838306/c
LOCUS AV838306 32 bp mRNA linear EST 07-NOV-2001
DEFINITION AV838306 Nori Satoh unpublished cDNA library, egg Ciona intestinalis cDNA clone rcieg03c03, mRNA sequence.
ACCESSION AV838306
VERSION AV838306.1 GI:16782457
KEYWORDS EST.
SOURCE Ciona intestinalis.
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 32)
AUTHORS Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
TITLE Expressed genes in Ciona intestinalis
JOURNAL Unpublished (2000)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES Location/Qualifiers
source 1. .32
/organism="Ciona intestinalis"
/db_xref="taxon:7719"

```
                     /clone="rcieg03c03"
                     /clone_lib="Nori Satoh unpublished cDNA library, egg"
                     /tissue_type="whole animal"
                     /dev_stage="egg"
                     /note="Vector: pBluescript SK"
BASE COUNT       13 a      4 c      5 g      9 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              7.92e+03        Length:         32
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     10              Gaps:           0

US-09-726-470A-2 (1-8) x AV838306 (1-32)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     23 CGGACTTTATCTTTT 9

RESULT 4
BH792463/c
LOCUS       BH792463                 33 bp    DNA     linear   GSS 02-APR-2002
DEFINITION  SALK_064289.24.00.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_064289.24.00.x, DNA
            sequence.
ACCESSION   BH792463
VERSION     BH792463.1  GI:19889261
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .33
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_064289.24.00.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
BASE COUNT       10 a      4 c      6 g     13 t
ORIGIN

Alignment Scores:
Pred. No.:              8.24e+03        Length:         33
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     17              Gaps:           0

US-09-726-470A-2 (1-8) x BH792463 (1-33)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     19 CGCACACTCAGTTTC 5

RESULT 5
BH792464/c
LOCUS       BH792464                 33 bp    DNA     linear   GSS 02-APR-2002
DEFINITION  SALK_064290.23.20.x Arabidopsis thaliana TDNA insertion lines
            Arabidopsis thaliana genomic clone SALK_064290.23.20.x, DNA
            sequence.
ACCESSION   BH792464
VERSION     BH792464.1  GI:19889263
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .33
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_064290.23.20.x"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
BASE COUNT       10 a      4 c      6 g     13 t
ORIGIN

Alignment Scores:
Pred. No.:              8.24e+03        Length:         33
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     17              Gaps:           0

US-09-726-470A-2 (1-8) x BH792464 (1-33)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     19 CGCACACTCAGTTTC 5

RESULT 6
AQ024967
```

```
LOCUS       AQ024967                42 bp    DNA     linear   GSS 23-AUG-2000
DEFINITION  EP(2)0827 Drosophila melanogaster EP line Drosophila melanogaster
            genomic Sequence recovered from 5' end of P element, DNA sequence.
ACCESSION   AQ024967
VERSION     AQ024967.1  GI:3265319
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 42)
  AUTHORS   Liao,G.-C., Rehm,E.J. and Rubin,G.M.
  TITLE     Insertion site preferences of the P transposable element in
            Drosophila melanogaster
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3347-3351 (2000)
  MEDLINE   20202638
COMMENT     Contact: Gerald Rubin
            Berkeley Drosophila Genome Project
            University of California, Berkeley
            LSA Building, Berkeley, CA 94720-3200, USA
            Fax: 5106439947
            Email: gerry@fruitfly.berkeley.edu
            Sequence recovery method was inverse PCR.

            Sequence orientation is forward strand relative to 5' end of P
            element

            The P element insertion position is base 35 in the 42 bases. This
            insertion position refers to the first base of the 8 base target
            recognition sequence.
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .42
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="Drosophila melanogaster EP line"
                     /note="Inverse PCR was performed on Drosophila
                     melanogaster strains each of which contains a single EP
                     transposable element insertion. (The generation of these
                     insertion strains is described in Rorth P, Szabo K, Bailey
                     A, Laverty T, Rehm J, Rubin GM, Weigmann K, Milan M, Benes
                     V, Ansorge W, Cohen SM. 1998. Systematic gain-of-function
                     genetics in Drosophila. Development 6:1049-1057.) The
                     resultant fragment for each strain was directly sequenced
                     to determine the genomic sequence at the site of
                     insertion. Details of the protocols used can be found at
                     http://fruitfly.berkeley.edu/p_disrupt/inverse_pcr.html."
BASE COUNT        6 a     16 c      7 g     13 t
ORIGIN

Alignment Scores:
Pred. No.:              1.12e+04        Length:         42
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     17              Gaps:           0

US-09-726-470A-2 (1-8) x AQ024967 (1-42)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      2 CGCTCTCTTTCCTTT 16

RESULT 7
AA960099/c
LOCUS       AA960099                43 bp    mRNA    linear   EST 08-MAY-1998
DEFINITION  ub54b07.s1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:1381525 3' similar to TR:Q26544 Q26544 G PROTEIN BETA
            SUBUNIT-LIKE PROTEIN TRANS-SPLICED. ;, mRNA sequence.
ACCESSION   AA960099
VERSION     AA960099.1  GI:3125999
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 43)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:903993
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .43
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1381525"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       13 a      6 c     17 g      7 t
ORIGIN

Alignment Scores:
Pred. No.:              1.16e+04        Length:         43
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     9               Gaps:           0

US-09-726-470A-2 (1-8) x AA960099 (1-43)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     19 CGCTCCTTGACCTTT 5

RESULT 8
AI316449/c
LOCUS       AI316449                43 bp    mRNA    linear   EST 17-DEC-1998
DEFINITION  uj60d11.y1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1924341 5' similar to SW:SP24_RAT Q62740 SECRETED
            PHOSPHOPROTEIN 24 ;, mRNA sequence.
ACCESSION   AI316449
VERSION     AI316449.1  GI:4031716
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 43)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:980633
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .43
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1924341"
                     /clone_lib="Sugano mouse liver mlia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT        9 a     12 c     13 g      9 t
ORIGIN

Alignment Scores:
Pred. No.:              1.16e+04        Length:         43
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     9               Gaps:           0

US-09-726-470A-2 (1-8) x AI316449 (1-43)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     42 AGGACGTTGACTTTC 28

RESULT 9
AQ025720
LOCUS       AQ025720                 44 bp    DNA     linear   GSS 30-JUN-1998
DEFINITION  l(2)k03201 Drosophila melanogaster P lethal line Drosophila
            melanogaster genomic Sequence recovered from 5' end of P element,
            DNA sequence.
ACCESSION   AQ025720
VERSION     AQ025720.1  GI:3266072
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 44)
  AUTHORS   Spradling,A.C., Stern,D., Beaton,A., Rehm,E.J., Laverty,T., Mozden
            ,N., Misra,S. and Rubin,G.M.
  TITLE     The BDGP gene disruption project: Single P element insertions
            mutating 30% of Drosophila autosomal genes
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Gerald Rubin
            Berkeley Drosophila Genome Project
            University of California, Berkeley
            LSA Building, Berkeley, CA 94720-3200, USA
            Fax: 5106439947
            Email: gerry@fruitfly.berkeley.edu
            Sequence recovery method was inverse PCR.

            Sequence orientation is forward strand relative to 5' end of P
            element

            The P element insertion position is base 037 in the 44 bases. This
            insertion position refers to the first base of the 8 base target
            recognition sequence.
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .44
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="Drosophila melanogaster P lethal line"
                     /note="Inverse PCR was performed on Drosophila
                     melanogaster strains each of which contains a single P
                     transposable element insertion that is thought to cause
                     either lethality or sterility. The resultant fragment for
                     each strain was directly sequenced to determine the
                     genomic sequence at the site of insertion. Details of the
                     protocols used can be found at
                     http://fruitfly.berkeley.edu/p_disrupt/inverse_pcr.html."
BASE COUNT        8 a     12 c     11 g     13 t
ORIGIN

Alignment Scores:
Pred. No.:              1.19e+04        Length:         44
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     17              Gaps:           0

US-09-726-470A-2 (1-8) x AQ025720 (1-44)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     16 CGAAGTTTGACATTT 30

RESULT 10
AQ025751
LOCUS       AQ025751                 45 bp    DNA     linear   GSS 30-JUN-1998
DEFINITION  l(2)k04917 Drosophila melanogaster P lethal line Drosophila
            melanogaster genomic Sequence recovered from 5' end of P element,
            DNA sequence.
ACCESSION   AQ025751
VERSION     AQ025751.1  GI:3266103
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   Spradling,A.C., Stern,D., Beaton,A., Rehm,E.J., Laverty,T., Mozden
            ,N., Misra,S. and Rubin,G.M.
  TITLE     The BDGP gene disruption project: Single P element insertions
```

```
            mutating 30% of Drosophila autosomal genes
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Gerald Rubin
            Berkeley Drosophila Genome Project
            University of California, Berkeley
            LSA Building, Berkeley, CA 94720-3200, USA
            Fax: 5106439947
            Email: gerry@fruitfly.berkeley.edu
            Sequence recovery method was inverse PCR.

            Sequence orientation is forward strand relative to 5' end of P
            element

            The P element insertion position is base 038 in the 45 bases. This
            insertion position refers to the first base of the 8 base target
            recognition sequence.
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="Drosophila melanogaster P lethal line"
                     /note="Inverse PCR was performed on Drosophila
                     melanogaster strains each of which contains a single P
                     transposable element insertion that is thought to cause
                     either lethality or sterility. The resultant fragment for
                     each strain was directly sequenced to determine the
                     genomic sequence at the site of insertion. Details of the
                     protocols used can be found at
                     http://fruitfly.berkeley.edu/p_disrupt/inverse_pcr.html."
BASE COUNT        8 a     13 c     11 g     13 t
ORIGIN

Alignment Scores:
Pred. No.:              1.23e+04        Length:         45
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     17              Gaps:           0

US-09-726-470A-2 (1-8) x AQ025751 (1-45)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     16 CGAAGTTTGACATTT 30

RESULT 11
AQ025803
LOCUS       AQ025803                47 bp    DNA     linear   GSS 30-JUN-1998
DEFINITION  l(2)k06617 Drosophila melanogaster P lethal line Drosophila
            melanogaster genomic Sequence recovered from 5' end of P element,
            DNA sequence.
ACCESSION   AQ025803
VERSION     AQ025803.1  GI:3266155
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 47)
  AUTHORS   Spradling,A.C., Stern,D., Beaton,A., Rehm,E.J., Laverty,T., Mozden
            ,N., Misra,S. and Rubin,G.M.
  TITLE     The BDGP gene disruption project: Single P element insertions
            mutating 30% of Drosophila autosomal genes
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Gerald Rubin
            Berkeley Drosophila Genome Project
            University of California, Berkeley
            LSA Building, Berkeley, CA 94720-3200, USA
            Fax: 5106439947
            Email: gerry@fruitfly.berkeley.edu
            Sequence recovery method was inverse PCR.

            Sequence orientation is forward strand relative to 5' end of P
            element

            The P element insertion position is base 040 in the 47 bases. This
            insertion position refers to the first base of the 8 base target
            recognition sequence.
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .47
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="Drosophila melanogaster P lethal line"
                     /note="Inverse PCR was performed on Drosophila
                     melanogaster strains each of which contains a single P
                     transposable element insertion that is thought to cause
                     either lethality or sterility. The resultant fragment for
                     each strain was directly sequenced to determine the
                     genomic sequence at the site of insertion. Details of the
                     protocols used can be found at
                     http://fruitfly.berkeley.edu/p_disrupt/inverse_pcr.html."
BASE COUNT        8 a     19 c      7 g     13 t
ORIGIN

Alignment Scores:
Pred. No.:              1.3e+04         Length:         47
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     17              Gaps:           0

US-09-726-470A-2 (1-8) x AQ025803 (1-47)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      2 CGCTCTCTTTCCTTT 16

RESULT 12
AZ499421/c
LOCUS       AZ499421                48 bp    DNA     linear   GSS 05-OCT-2000
DEFINITION  1M0337I12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0337I12 F, DNA sequence.
ACCESSION   AZ499421
VERSION     AZ499421.1  GI:10678231
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 48)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0337  row: I  column: 12
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 48.
FEATURES             Location/Qualifiers
```

```
     source          1. .48
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0337I12"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       24 a      5 c      6 g     13 t
ORIGIN

Alignment Scores:
Pred. No.:                1.33e+04        Length:         48
Score:                    15.00           Matches:        3
Percent Similarity:       60.00%          Conservative:   0
Best Local Similarity:    60.00%          Mismatches:     2
Query Match:              75.00%          Indels:         0
DB:                       17              Gaps:           0

US-09-726-470A-2 (1-8) x AZ499421 (1-48)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     38 CGTTCACTTTCTTTT 24

RESULT 13
AU105436
LOCUS       AU105436                 50 bp    mRNA    linear   EST 30-AUG-2001
DEFINITION  AU105436 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            HEP15082, mRNA sequence.
ACCESSION   AU105436
VERSION     AU105436.1  GI:13554957
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki
            ,Y., Nakamura,Y., Suyama,A. and Sugano,S.
  TITLE     Diverse transcriptional initiation revealed by fine, large-scale
            mapping of mRNA start sites
  JOURNAL   EMBO Rep. 2 (5), 388-393 (2001)
  MEDLINE   21270072
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES              Location/Qualifiers
     source           1. .50
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="HEP15082"
                      /clone_lib="Sugano Homo sapiens cDNA library"
                      /note="Differential display comparison of untreated and
                      dimethylfumarate treated U937 cells"
BASE COUNT        3 a      23 c      8 g     16 t
ORIGIN

Alignment Scores:
Pred. No.:                1.41e+04        Length:         50
Score:                    15.00           Matches:        3
Percent Similarity:       60.00%          Conservative:   0
Best Local Similarity:    60.00%          Mismatches:     2
Query Match:              75.00%          Indels:         0
DB:                       9               Gaps:           0

US-09-726-470A-2 (1-8) x AU105436 (1-50)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     12 CGCTCCCTTACATTT 26

RESULT 14
AU105781
LOCUS       AU105781                 50 bp    mRNA    linear   EST 30-AUG-2001
DEFINITION  AU105781 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            ZRV61136, mRNA sequence.
ACCESSION   AU105781
VERSION     AU105781.1  GI:13555302
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki
            ,Y., Nakamura,Y., Suyama,A. and Sugano,S.
  TITLE     Diverse transcriptional initiation revealed by fine, large-scale
            mapping of mRNA start sites
  JOURNAL   EMBO Rep. 2 (5), 388-393 (2001)
  MEDLINE   21270072
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES              Location/Qualifiers
     source           1. .50
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="ZRV61136"
                      /clone_lib="Sugano Homo sapiens cDNA library"
                      /note="Differential display comparison of untreated and
                      dimethylfumarate treated U937 cells"
BASE COUNT        9 a      13 c      15 g     13 t
ORIGIN

Alignment Scores:
Pred. No.:                1.41e+04        Length:         50
Score:                    15.00           Matches:        3
Percent Similarity:       60.00%          Conservative:   0
Best Local Similarity:    60.00%          Mismatches:     2
Query Match:              75.00%          Indels:         0
DB:                       9               Gaps:           0

US-09-726-470A-2 (1-8) x AU105781 (1-50)
```

```
Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     24 CGAGCCTTAGCTTTC 38

RESULT 15
AU105834
LOCUS       AU105834                  50 bp    mRNA    linear   EST 30-AUG-2001
DEFINITION  AU105834 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            KAT11187, mRNA sequence.
ACCESSION   AU105834
VERSION     AU105834.1  GI:13555355
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K., Sakaki
            ,Y., Nakamura,Y., Suyama,A. and Sugano,S.
  TITLE     Diverse transcriptional initiation revealed by fine, large-scale
            mapping of mRNA start sites
  JOURNAL   EMBO Rep. 2 (5), 388-393 (2001)
  MEDLINE   21270072
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="KAT11187"
                     /clone_lib="Sugano Homo sapiens cDNA library"
                     /note="Differential display comparison of untreated and
                     dimethylfumarate treated U937 cells"
BASE COUNT        9 a     12 c     14 g     15 t
ORIGIN

Alignment Scores:
Pred. No.:              1.41e+04        Length:         50
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     9               Gaps:           0

US-09-726-470A-2 (1-8) x AU105834 (1-50)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     27 CGGTCGCTGTCTTTC 41


Search completed: December 14, 2002, 17:45:54
Job time : 1556 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: December 14, 2002, 15:41:54 ; Search time 57.5 Seconds
(without alignments)
28.667 Million cell updates/sec

Title: US-09-726-470A-2
Perfect score: 20
Sequence: 1 XXXRXLXF 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	15	75.0	39	16	Q8X417	Q8x417 escherichia
2	15	75.0	42	8	Q33005	Q33005 pinus thunb
3	15	75.0	46	5	Q8WR26	Q8wr26 anopheles g
4	15	75.0	47	4	Q96EU3	Q96eu3 homo sapien
5	15	75.0	53	2	Q9RAW5	Q9raw5 frankia sp.
6	15	75.0	60	16	Q8X8T7	Q8x8t7 escherichia
7	15	75.0	63	16	Q929V4	Q929v4 listeria in
8	15	75.0	63	16	Q8Y5J7	Q8y5j7 listeria mo
9	15	75.0	64	12	O71115	O71115 trichoplusi
10	15	75.0	68	8	Q8SFT9	Q8sft9 homarus gam
11	15	75.0	69	14	Q99IS8	Q99is8 uncultured
12	15	75.0	75	14	Q99IR7	Q99ir7 uncultured
13	15	75.0	76	2	Q9L8X7	Q918x7 enterococcu
14	15	75.0	77	2	Q54148	Q54148 shigella fl
15	15	75.0	77	2	Q8VSD1	Q8vsd1 shigella fl
16	15	75.0	78	10	Q942F4	Q942f4 oryza sativ
17	15	75.0	78	12	Q91MW1	Q91mw1 lumpy skin
18	15	75.0	78	16	Q98G90	Q98g90 rhizobium l
19	15	75.0	78	16	Q8XPZ2	Q8xpz2 ralstonia s
20	15	75.0	79	16	Q9I2D4	Q9i2d4 pseudomonas
21	15	75.0	80	3	O14280	O14280 schizosacch
22	15	75.0	90	12	Q85345	Q85345 vaccinia vi
23	15	75.0	93	16	Q8ZQT8	Q8zqt8 salmonella
24	15	75.0	93	16	Q8Z8C3	Q8z8c3 salmonella
25	15	75.0	95	2	O32936	O32936 mycobacteri
26	15	75.0	95	10	Q8S019	Q8s019 oryza sativ
27	15	75.0	95	16	Q9CM66	Q9cm66 pasteurella
28	15	75.0	97	17	Q8TRU8	Q8tru8 methanosarc
29	15	75.0	98	16	Q981R5	Q981r5 rhizobium l
30	15	75.0	98	16	Q92ZE9	Q92ze9 rhizobium m
31	15	75.0	100	16	O31627	O31627 bacillus su
32	15	75.0	101	16	Q928A0	Q928a0 listeria in
33	15	75.0	101	16	Q8Y4F1	Q8y4f1 listeria mo
34	15	75.0	102	2	Q93KR2	Q93kr2 yersinia en
35	15	75.0	103	17	Q8TWT6	Q8twt6 methanopyru
36	15	75.0	104	12	Q8VBF4	Q8vbf4 white spot
37	15	75.0	105	8	O99903	O99903 mugil cepha
38	15	75.0	105	11	Q9DB71	Q9db71 mus musculu
39	15	75.0	105	15	Q9PXX6	Q9pxx6 bovine leuk
40	15	75.0	107	8	O79295	O79295 scolopax mi
41	15	75.0	108	2	Q9LA15	Q9la15 thiobacillu
42	15	75.0	109	16	O25421	O25421 helicobacte
43	15	75.0	111	8	Q9MTN3	Q9mtn3 oenothera h
44	15	75.0	112	12	Q91LK8	Q91lk8 white spot
45	15	75.0	113	8	Q9G3Z4	Q9g3z4 rachycentro

ALIGNMENTS

RESULT 1
Q8X417
ID Q8X417 PRELIMINARY; PRT; 39 AA.
AC Q8X417;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein z4614.
GN Z4614.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
DR EMBL; AE005553; AAG58382.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 39 AA; 4604 MW; 9607DF8C26905A1D CRC64;

Query Match 75.0%; Score 15; DB 16; Length 39;
Best Local Similarity 60.0%; Pred. No. 4.3e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       4 RXLXF 8
           | | |
Db      28 RALAF 32
```

RESULT 2

```
Q33005
ID   Q33005      PRELIMINARY;      PRT;    42 AA.
AC   Q33005;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF42g.
OS   Pinus thunbergii (Green pine) (Japanese black pine).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX   NCBI_TaxID=3350;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92212283; PubMed=1557027;
RA   Tsudzuki J., Nakashima K., Tsudzuki T., Hiratsuka J., Shibata M.,
RA   Wakasugi T., Sugiura M.;
RT   "Chloroplast DNA of black pine retains a residual inverted repeat
RT   lacking rRNA genes: nucleotide sequences of trnQ, trnK, psbA, trnI and
RT   trnH and the absence of rps16.";
RL   Mol. Gen. Genet. 232:206-214(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95094312; PubMed=8001170;
RA   Tsudzuki J., Ito S., Tsudzuki T., Wakasugi T., Sugiura M.;
RT   "A new gene encoding tRNA pro (GGG) is present in the chloroplast
RT   genome of black pine: a compilation of 32 tRNA genes from black pine
RT   chloroplasts.";
RL   Curr. Genet. 26:153-158(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95024047; PubMed=7937893;
RA   Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA   Sugiura M.;
RT   "Loss of all ndh genes as determined by sequencing the entire
RT   chloroplast genome of the black pine Pinus thunbergii.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
DR   EMBL; D17510; BAA04457.1; -.
KW   Chloroplast.
SQ   SEQUENCE   42 AA;  4945 MW;  932D81DDA0604964 CRC64;

  Query Match             75.0%;  Score 15;  DB 8;  Length 42;
  Best Local Similarity   60.0%;  Pred. No. 4.6e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db      14 RSLSF 18


RESULT 3
Q8WR26
ID   Q8WR26      PRELIMINARY;      PRT;    46 AA.
AC   Q8WR26;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical 5.3 kDa protein.
OS   Anopheles gambiae (African malaria mosquito).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC   Anopheles.
OX   NCBI_TaxID=7165;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Francischetti I.M., Valenzuela J.G., Ribeiro J.M.;
RT   "Towards a catalog for genes and proteins from the salivary gland of
RT   the malaria vector, Anopheles gambiae.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF457561; AAL68791.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   46 AA;  5266 MW;  9D378CAAF35E4CD2 CRC64;

  Query Match             75.0%;  Score 15;  DB 5;  Length 46;
  Best Local Similarity   60.0%;  Pred. No. 5e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db      16 RALSF 20


RESULT 4
Q96EU3
ID   Q96EU3      PRELIMINARY;      PRT;    47 AA.
AC   Q96EU3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Unknown (protein for MGC:20802).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=UTERUS;
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC011940; AAH11940.1; -.
SQ   SEQUENCE   47 AA;  5268 MW;  F7B598AB9A649581 CRC64;

  Query Match             75.0%;  Score 15;  DB 4;  Length 47;
  Best Local Similarity   60.0%;  Pred. No. 5.1e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db      25 RSLAF 29


RESULT 5
Q9RAW5
ID   Q9RAW5      PRELIMINARY;      PRT;    53 AA.
AC   Q9RAW5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Putative dienelactone hydrolase-like protein (Fragment).
OS   Frankia sp. ACN14a/ts-r.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Frankineae; Frankiaceae; Frankia.
OX   NCBI_TaxID=92643;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ACN14A/TS-R;
RX   MEDLINE=20117996; PubMed=10652097;
RA   Pouch M.-N., Cournoyer B., Baumeister W.;
RT   "Characterization of the 20S proteasome from the actinomycete
RT   Frankia.";
RL   Mol. Microbiol. 35:368-377(2000).
DR   EMBL; AF142435; AAF14269.1; -.
DR   InterPro; IPR002925; DLH.
DR   Pfam; PF01738; DLH; 1.
KW   Hydrolase.
FT   NON_TER       1      1
SQ   SEQUENCE   53 AA;  5748 MW;  CAF445055B8B1E87 CRC64;

  Query Match             75.0%;  Score 15;  DB 2;  Length 53;
  Best Local Similarity   60.0%;  Pred. No. 5.7e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db      41 RALAF 45
```

```
RESULT 6
Q8X8T7
ID   Q8X8T7      PRELIMINARY;      PRT;    60 AA.
AC   Q8X8T7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein z2370.
GN   Z2370 OR ECS2740.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AE005369; AAG56416.1; -.
DR   EMBL; AP002559; BAB36163.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   60 AA;  6544 MW;  AE13AA97B7255109 CRC64;

  Query Match             75.0%;  Score 15;  DB 16;  Length 60;
  Best Local Similarity   60.0%;  Pred. No. 6.4e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db      29 RALAF 33


RESULT 7
Q929V4
ID   Q929V4      PRELIMINARY;      PRT;    63 AA.
AC   Q929V4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein lin2169.
GN   LIN2169.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Listeriaceae; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / SEROVAR 6A;
RX   PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL596171; CAC97399.1; -.
DR   ListiList; LIN02169; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   63 AA;  7418 MW;  B7A15C4291E86115 CRC64;

  Query Match             75.0%;  Score 15;  DB 16;  Length 63;
  Best Local Similarity   60.0%;  Pred. No. 6.7e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db      39 RTLTF 43


RESULT 8
Q8Y5J7
ID   Q8Y5J7      PRELIMINARY;      PRT;    63 AA.
AC   Q8Y5J7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein lmo2063.
GN   LMO2063.
OS   Listeria monocytogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Listeriaceae; Listeria.
OX   NCBI_TaxID=1639;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGD-E / SEROVAR 1/2A;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL591982; CAD00141.1; -.
DR   ListiList; LMO02063; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   63 AA;  7504 MW;  A31E07098B5D3050 CRC64;

  Query Match             75.0%;  Score 15;  DB 16;  Length 63;
  Best Local Similarity   60.0%;  Pred. No. 6.7e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db      39 RTLTF 43


RESULT 9
O71115
ID   O71115      PRELIMINARY;      PRT;    64 AA.
AC   O71115;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
```

```
DE   P7.41.
OS   Trichoplusia ni granulosis virus (TnGV) (Trichoplusia ni
OS   granulovirus).
OC   Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
OX   NCBI_TaxID=10462;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98264509; PubMed=9603347;
RA   Bideshi D.K., Hice R.H., Ge B., Federici B.A.;
RT   "Molecular characterization and expression of the Trichoplusia ni
RT   granulovirus helicase gene.";
RL   J. Gen. Virol. 79:1309-1319(1998).
DR   EMBL; AF032994; AAC40853.1; -.
SQ   SEQUENCE   64 AA;  7416 MW;  AA541DB3DDC74ED6 CRC64;

  Query Match             75.0%;  Score 15;  DB 12;  Length 64;
  Best Local Similarity   60.0%;  Pred. No. 6.8e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db     15 RALSF 19


RESULT 10
Q8SFT9
ID   Q8SFT9      PRELIMINARY;      PRT;    68 AA.
AC   Q8SFT9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cytochrome b (Fragment).
GN   CYTB.
OS   Homarus gammarus (European lobster) (Homarus vulgaris).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC   Nephropoidea; Nephropidae; Homarus.
OX   NCBI_TaxID=6707;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Katsares V., Apostolidis A., Triantafyllidis A., Kouvatsi A.,
RA   Triantaphyllidis C.;
RT   "Development of mitochondrial primers for use with homarid lobster.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF494203; AAM15924.1; -.
KW   Mitochondrion.
FT   NON_TER       1      1
SQ   SEQUENCE   68 AA;  8074 MW;  99EFF029E09DC1D0 CRC64;

  Query Match             75.0%;  Score 15;  DB 8;  Length 68;
  Best Local Similarity   60.0%;  Pred. No. 7.2e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db      3 RSLTF 7


RESULT 11
Q99IS8
ID   Q99IS8      PRELIMINARY;      PRT;    69 AA.
AC   Q99IS8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Hypothetical 7.7 kDa protein.
OS   uncultured organism.
OC   unclassified; environmental samples.
OX   NCBI_TaxID=155900;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stokes H.W., Nield B.S., Mabbutt B.C., Nevalainen H., Holmes A.J.,
RA   Gillings M.R.;
RT   "Novel and diverse integron-like gene cassettes are prevalent in
RT   natural environments.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF349105; AAK28612.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   69 AA;  7736 MW;  EB73269523C2B349 CRC64;

  Query Match             75.0%;  Score 15;  DB 14;  Length 69;
  Best Local Similarity   60.0%;  Pred. No. 7.3e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db     21 RALSF 25


RESULT 12
Q99IR7
ID   Q99IR7      PRELIMINARY;      PRT;    75 AA.
AC   Q99IR7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Hypothetical 8.3 kDa protein.
OS   uncultured organism.
OC   unclassified; environmental samples.
OX   NCBI_TaxID=155900;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stokes H.W., Nield B.S., Mabbutt B.C., Nevalainen H., Holmes A.J.,
RA   Gillings M.R.;
RT   "Novel and diverse integron-like gene cassettes are prevalent in
RT   natural environments.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF349110; AAK28623.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   75 AA;  8285 MW;  B0DAC6D7E3CF39C2 CRC64;

  Query Match             75.0%;  Score 15;  DB 14;  Length 75;
  Best Local Similarity   60.0%;  Pred. No. 7.8e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db     27 RALSF 31


RESULT 13
Q9L8X7
ID   Q9L8X7      PRELIMINARY;      PRT;    76 AA.
AC   Q9L8X7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   Hypothetical 8.7 kDa protein.
OS   Enterococcus faecalis (Streptococcus faecalis).
OG   Plasmid pIP834.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC   Enterococcaceae; Enterococcus.
OX   NCBI_TaxID=1351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BM4382; TRANSPOSON=Tn1549;
RA   Garnier F., Taourit S., Glaser P., Courvalin P., Galimand M.;
RT   "Characterization of transposon Tn1549 conferring VanB-type resistance
RT   in Enterococcus sp.";
RL   Microbiology 0:0-0(2000).
DR   EMBL; AF192329; AAF72366.1; -.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   76 AA;  8698 MW;  C628291E1C80D050 CRC64;

  Query Match             75.0%;  Score 15;  DB 2;  Length 76;
```

```
  Best Local Similarity   60.0%;  Pred. No. 7.9e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db       7 RALSF 11


RESULT 14
Q54148
ID   Q54148      PRELIMINARY;      PRT;    77 AA.
AC   Q54148;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 8.9 kDa protein.
GN   S0089.
OS   Shigella flexneri.
OG   Plasmid pWR100, and Plasmid virulence pWR501.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Shigella.
OX   NCBI_TaxID=623;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M90T-W / SEROTYPE 5; PLASMID=PWR100;
RX   MEDLINE=92167809; PubMed=1791758;
RA   Venkatesan M.M., Buysse J.M., Hartman A.B.;
RT   "Sequence variation in two ipaH genes of Shigella flexneri 5 and
RT   homology to the LRG-like family of proteins.";
RL   Mol. Microbiol. 5:2435-2445(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M90T-W / SEROTYPE 5; PLASMID=PWR100;
RX   MEDLINE=97074644; PubMed=8917071;
RA   Venkatesan M.M., Alexander W.A., Fernandez-Prada C.;
RT   "A Shigella flexneri invasion plasmid gene, ipgH, with homology to
RT   IS629 and sequences encoding bacterial sugar phosphate transport
RT   proteins.";
RL   Gene 175:23-27(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   PLASMID=VIRULENCE PWR501;
RA   Venkatesan M.M., Goldberg M.B., Rose D.J., Grotbeck E.J., Burland V.,
RA   Blattner F.R.;
RT   "Complete DNA Sequence and Analysis of the Large Virulence Plasmid of
RT   Shigella flexneri.";
RL   Infect. Immun. 0:0-0(2001).
DR   EMBL; U28354; AAC44576.1; -.
DR   EMBL; AF348706; AAK18399.1; -.
DR   InterPro; IPR002611; IstB_ATPbind.
DR   Pfam; PF01695; IstB; 1.
KW   Plasmid.
SQ   SEQUENCE   77 AA;  8875 MW;  01C1DFDA949974C3 CRC64;

  Query Match            75.0%;  Score 15;  DB 2;  Length 77;
  Best Local Similarity   60.0%;  Pred. No. 8e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db      49 RSLSF 53


RESULT 15
Q8VSD1
ID   Q8VSD1      PRELIMINARY;      PRT;    77 AA.
AC   Q8VSD1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 8.9 kDa protein.
GN   CP0214.
OS   Shigella flexneri 2a.
OG   Plasmid pCP301.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Shigella.
OX   NCBI_TaxID=42897;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=301;
RA   Jin Q., Zhang J.Y., Liu H., Yang J., Yang F., Zhang X.B., Wang J.H.,
RA   Yang G.W., Wu H.T., Dong J., Sun L.L., Xue Y., Zhao A.L., Gao Y.S.,
RA   Zhu J.P., Kan B., Chen S.X., Yao Z.J., He B.K., Chen R.S., Ma D.L.,
RA   Yuan Z.H., Xu J.G., Wang Y., Shen Y., Lu W.C., Qiang B.Q., Wen Y.M.,
RA   Hou Y.D.;
RT   "Complete DNA sequence and analysis of the large virulence plasmid
RT   pCP301 of Shigella flexneri.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF386526; AAL72411.1; -.
DR   InterPro; IPR002611; IstB_ATPbind.
DR   Pfam; PF01695; IstB; 1.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   77 AA;  8860 MW;  01D2426C5C696EEA CRC64;

  Query Match            75.0%;  Score 15;  DB 2;  Length 77;
  Best Local Similarity   60.0%;  Pred. No. 8e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db      49 RSLSF 53


Search completed: December 14, 2002, 15:48:52
Job time : 60.5 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: December 14, 2002, 15:34:14 ; Search time 23.5 Seconds
(without alignments)
14.120 Million cell updates/sec

Title: US-09-726-470A-2
Perfect score: 20
Sequence: 1 XXXRXLXF 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	15	75.0	71	1	Y4UG_RHISN	P55671 rhizobium s
2	15	75.0	79	1	CATR_HUMAN	Q13166 homo sapien
3	15	75.0	97	1	YBGE_ECOLI	P37343 escherichia
4	15	75.0	102	1	RS24_AERPE	Q9ycy0 aeropyrum p
5	15	75.0	104	1	URE2_MYCTU	P50048 mycobacteri
6	15	75.0	110	1	YFMT_THETH	P43520 thermus the
7	15	75.0	111	1	YCX5_OENHO	Q9mtn3 oenothera h
8	15	75.0	116	1	PHS_XYLFA	Q9pab4 xylella fas
9	15	75.0	119	1	AMCY_METEX	P04172 methylobact
10	15	75.0	121	1	HV2E_HUMAN	P01818 homo sapien
11	15	75.0	129	1	NRDJ_BACSU	O31876 bacillus su
12	15	75.0	133	1	LECA_ARTIN	P18670 artocarpus
13	15	75.0	133	1	LECA_MACPO	P18674 maclura pom
14	15	75.0	141	1	V223_FOWPV	Q9j512 fowlpox vir
15	15	75.0	141	1	YEF5_YEAST	P32616 saccharomyc
16	15	75.0	146	1	Y677_HAEIN	P44036 haemophilus
17	15	75.0	155	1	GST1_HUMAN	P10620 homo sapien
18	15	75.0	157	1	IBP_BUCAI	P57640 buchnera ap
19	15	75.0	161	1	Y311_RICPR	Q9zdl6 rickettsia
20	15	75.0	189	1	INA4_HUMAN	P05014 homo sapien
21	15	75.0	189	1	INAA_HUMAN	P01566 homo sapien
22	15	75.0	189	1	INAD_HUMAN	P01570 homo sapien
23	15	75.0	189	1	INAG_HUMAN	P01571 homo sapien
24	15	75.0	192	1	OASB_MOUSE	Q60856 mus musculu
25	15	75.0	193	1	RNFA_ECOLI	P76181 escherichia
26	15	75.0	197	1	HIS7_CLOAB	Q97ki1 clostridium
27	15	75.0	204	1	COAE_RALSO	Q8xvk2 ralstonia s
28	15	75.0	208	1	RL4_MYCCA	P10135 mycoplasma
29	15	75.0	211	1	WFD1_MOUSE	Q9esh5 mus musculu
30	15	75.0	227	1	YPVA_METTF	P29577 methanobact
31	15	75.0	228	1	VHEL_LSV	P27330 lily sympto
32	15	75.0	233	1	LICC_HAEIN	P14183 haemophilus
33	15	75.0	237	1	PYRF_LACPL	P77888 lactobacill
34	15	75.0	238	1	EPCR_HUMAN	Q9unn8 homo sapien
35	15	75.0	239	1	ATP6_ENTHR	P43454 enterococcu
36	15	75.0	253	1	YF88_HELPJ	Q9zj24 helicobacte
37	15	75.0	253	1	YF88_HELPY	O26107 helicobacte
38	15	75.0	260	1	MURI_CLOPE	Q8xhb7 clostridium
39	15	75.0	265	1	HIS4_XYLFA	Q9pbc9 xylella fas
40	15	75.0	267	1	KORB_ARCFU	O29782 archaeoglob
41	15	75.0	276	1	APAH_NEIMA	Q9jvf4 neisseria m
42	15	75.0	277	1	DMA_VIBCH	Q08318 vibrio chol
43	15	75.0	281	1	Y149_MYCGE	P47395 mycoplasma
44	15	75.0	285	1	REP2_ECOLI	P03066 escherichia
45	15	75.0	286	1	HK52_MOUSE	P43687 mus musculu

ALIGNMENTS

RESULT 1
Y4UG_RHISN
ID Y4UG_RHISN STANDARD; PRT; 71 AA.
AC P55671;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 7.8 kDa protein Y4UG.
GN Y4UG.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96389014; PubMed=8796346;
RA Freiberg C., Perret X., Broughton W.J., Rosenthal A.;
RT "Sequencing the 500-kb GC-rich symbiotic replicon of Rhizobium sp.
RT NGR234 using dye terminators and a thermostable 'sequenase': a
RT beginning.";
RL Genome Res. 6:590-600(1996).
CC -!- SIMILARITY: NONE OBVIOUS.

DR EMBL; Z68203; -; NOT_ANNOTATED_CDS.
DR EMBL; AE000099; AAB91879.1; -.
KW Hypothetical protein; Transmembrane; Plasmid.
FT TRANSMEM 12 32 POTENTIAL.
SQ SEQUENCE 71 AA; 7769 MW; 655F2FDA41049001 CRC64;

Query Match 75.0%; Score 15; DB 1; Length 71;
Best Local Similarity 60.0%; Pred. No. 1e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      4 RXLXF 8
          | | |
Db      8 RSLSF 12
```

RESULT 2
CATR_HUMAN

```
ID   CATR_HUMAN     STANDARD;      PRT;    79 AA.
AC   Q13166;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   CATR tumorigenic conversion 1 protein (CATR1.3).
GN   CATR1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Carcinoma;
RX   MEDLINE=95327656; PubMed=7604004;
RA   Li D., Noyes I., Shuler C., Milo G.E.;
RT   "Cloning and sequencing of CATR1.3, a human gene associated with
RT   tumorigenic conversion.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:6409-6413(1995).
CC   -!- DEVELOPMENTAL STAGE: ASSOCIATED WITH TUMORIGENIC CONVERSION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U25433; -; NOT_ANNOTATED_CDS.
DR   Genew; HGNC:1525; CATR1.
DR   MIM; 600676; -.
SQ   SEQUENCE   79 AA;  9224 MW;  BC3667C059114CF3 CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 79;
  Best Local Similarity   60.0%;  Pred. No. 1.1e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db     42 RALTF 46


RESULT 3
YBGE_ECOLI
ID   YBGE_ECOLI     STANDARD;      PRT;    97 AA.
AC   P37343; P75755;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein ybgE.
GN   YBGE OR B0735 OR Z0903 OR ECS0770.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kim K., Allen E., Araujo R., Aparicio A.M., Botstein D.,
RA   Cherry M., Chung E., Dietrich F., Duncan M., Federspiel N.,
RA   Kalman S., Komp C., Lashkari D., Lew H., Lin D., Namath A.,
RA   Oefner P., Davis R.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
RN   [6]
RP   SEQUENCE OF 1-6 FROM N.A.
RX   MEDLINE=88330812; PubMed=2843510;
RA   Green G.N., Fang H., Lin R.-J., Newton G., Mather M., Georgiou C.D.,
RA   Gennis R.B.;
RT   "The nucleotide sequence of the cyd locus encoding the two subunits
RT   of the cytochrome d terminal oxidase complex of Escherichia coli.";
RL   J. Biol. Chem. 263:13138-13143(1988).
RN   [7]
RP   IDENTIFICATION.
RX   MEDLINE=95075659; PubMed=7984428;
RA   Borodovsky M., Rudd K.E., Koonin E.V.;
RT   "Intrinsic and extrinsic approaches for detecting genes in a
RT   bacterial genome.";
RL   Nucleic Acids Res. 22:4756-4767(1994).
RN   [8]
RP   IDENTIFICATION.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97221613; PubMed=9068659;
RA   Muller M.M., Webster R.E.;
RT   "Characterization of the tol-pal and cyd region of Escherichia coli
RT   K-12: transcript analysis and identification of two new proteins
RT   encoded by the cyd operon.";
RL   J. Bacteriol. 179:2077-2080(1997).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U30934; -; NOT_ANNOTATED_CDS.
DR   EMBL; AE000177; AAC73829.1; -.
DR   EMBL; D90713; BAA35401.1; -.
```

```
DR   EMBL; AE005252; AAG55071.1; -.
DR   EMBL; AP002553; BAB34193.1; -.
DR   EMBL; J03939; -; NOT_ANNOTATED_CDS.
DR   EcoGene; EG12395; ybgE.
KW   Complete proteome.
SQ   SEQUENCE   97 AA;  10932 MW;  2D34484BE99E8AC9 CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 97;
  Best Local Similarity   60.0%;  Pred. No. 1.4e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db     18 RALSF 22


RESULT 4
RS24_AERPE
ID   RS24_AERPE     STANDARD;      PRT;   102 AA.
AC   Q9YCY0;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   30S ribosomal protein S24e.
GN   RPS24E OR APE1132.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC   Desulfurococcaceae; Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
CC   -!- SIMILARITY: BELONGS TO THE S24E FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP000060; BAA80117.1; ALT_INIT.
DR   InterPro; IPR001976; Ribosomal_S24E.
DR   Pfam; PF01282; Ribosomal_S24e; 1.
DR   ProDom; PD006052; Ribosomal_S24E; 1.
DR   PROSITE; PS00529; RIBOSOMAL_S24E; FALSE_NEG.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   102 AA;  11858 MW;  DEAA205AAFBD8066 CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 102;
  Best Local Similarity   60.0%;  Pred. No. 1.4e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db     78 RALSF 82


RESULT 5
URE2_MYCTU
ID   URE2_MYCTU     STANDARD;      PRT;   104 AA.
AC   P50048;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Urease beta subunit (EC 3.5.1.5) (Urea amidohydrolase).
GN   UREB OR RV1849 OR MT1897 OR MTCY359.24C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=96004620; PubMed=7568014;
RA   Reyrat J.M., Berthet F.X., Gicquel B.;
RT   "The urease locus of Mycobacterium tuberculosis and its utilization
RT   for the demonstration of allelic exchange in Mycobacterium bovis
RT   bacillus Calmette-Guerin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:8768-8772(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Erdmann;
RX   MEDLINE=96032403; PubMed=7559354;
RA   Clemens D.L., Lee B.Y., Horwitz M.A.;
RT   "Purification, characterization, and genetic analysis of
RT   Mycobacterium tuberculosis urease, a potentially critical determinant
RT   of host-pathogen interaction.";
RL   J. Bacteriol. 177:5644-5652(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).
CC   -!- SUBUNIT: (ALPHA, BETA, GAMMA)(3) (BY SIMILARITY).
CC   -!- SIMILARITY: TO OTHER UREASES BETA SUBUNITS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L41141; AAC37006.1; -.
DR   EMBL; U33011; AAC43474.1; -.
DR   EMBL; Z83859; CAB06138.1; -.
DR   EMBL; AE007047; AAK46168.1; -.
DR   HSSP; P18315; 1FWB.
DR   TIGR; MT1897; -.
DR   TubercuList; Rv1849; -.
DR   InterPro; IPR002019; Urease_beta.
```

```
DR   Pfam; PF00699; Urease_beta; 1.
DR   ProDom; PD002326; Urease_beta; 1.
DR   TIGRFAMs; TIGR00192; urease_beta; 1.
KW   Hydrolase; Complete proteome.
SQ   SEQUENCE   104 AA;  11190 MW;  D621CE43A47304E0 CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 104;
  Best Local Similarity   60.0%;  Pred. No. 1.5e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db     48 RALSF 52


RESULT 6
YFMT_THETH
ID   YFMT_THETH     STANDARD;      PRT;   110 AA.
AC   P43520;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein in fmt 3'region (Fragment).
OS   Thermus thermophilus.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Thermales;
OC   Thermaceae; Thermus.
OX   NCBI_TaxID=274;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VK1;
RX   MEDLINE=95050326; PubMed=7961514;
RA   Meinnel T., Blanquet S.;
RT   "Characterization of the Thermus thermophilus locus encoding peptide
RT   deformylase and methionyl-tRNA(fMet) formyltransferase.";
RL   J. Bacteriol. 176:7387-7390(1994).
CC   -!- SIMILARITY: BELONGS TO THE UPF0042 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X79087; CAA55697.1; -.
DR   InterPro; IPR005337; UPF0042.
DR   Pfam; PF03668; UPF0042; 1.
KW   Hypothetical protein; ATP-binding.
FT   NP_BIND       8     15       ATP (POTENTIAL).
FT   NON_TER     110    110
SQ   SEQUENCE   110 AA;  12316 MW;  C4F21341300F9DA6 CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 110;
  Best Local Similarity   60.0%;  Pred. No. 1.6e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db     59 RALAF 63


RESULT 7
YCX5_OENHO
ID   YCX5_OENHO     STANDARD;      PRT;   111 AA.
AC   Q9MTN3;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical 12.8 kDa protein in ycf9-trnS intergenic region (ORF111).
OS   Oenothera hookeri (Hooker's evening primrose).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Myrtales; Onagraceae; Oenothera.
OX   NCBI_TaxID=85636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Johansen;
RX   MEDLINE=20309318; PubMed=10852478;
RA   Hupfer H., Swiatek M., Hornung S., Herrmann R.G., Maier R.M.,
RA   Chiu W.-L., Sears B.;
RT   "Complete nucleotide sequence of the Oenothera elata plastid
RT   chromosome, representing plastome I of the five distinguishable
RT   Euoenothera plastomes.";
RL   Mol. Gen. Genet. 263:581-585(2000).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ271079; CAB67142.1; -.
KW   Chloroplast; Hypothetical protein.
SQ   SEQUENCE   111 AA;  12814 MW;  E5E0CE989317F140 CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 111;
  Best Local Similarity   60.0%;  Pred. No. 1.6e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db      9 RALSF 13


RESULT 8
PHS_XYLFA
ID   PHS_XYLFA      STANDARD;      PRT;   116 AA.
AC   Q9PAB4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative pterin-4-alpha-carbinolamine dehydratase (EC 4.2.1.96) (PHS)
DE   (4-alpha-hydroxy-tetrahydropterin dehydratase) (Pterin carbinolamine
DE   dehydratase) (PCD).
GN   XF2604.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9a5c;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
```

```
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
CC   -!- CATALYTIC ACTIVITY: (6R)-6-(L-erythro-1,2-dihydroxypropyl)-
CC       5,6,7,8-tetrahydro-4a-hydroxypterin = (6R)-6-(L-erythro-1,2-
CC       dihydroxypropyl)-7,8-dihydro-6H-pterin + H(2)O.
CC   -!- SIMILARITY: BELONGS TO THE PTERIN-4-ALPHA-CARBINOLAMINE
CC       DEHYDRATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004067; AAF85401.1; -.
DR   InterPro; IPR001533; Trans_pterinDh.
DR   Pfam; PF01329; Pterin_4a; 1.
DR   ProDom; PD007262; Trans_pterinDh; 1.
KW   Hypothetical protein; Lyase; Complete proteome.
SQ   SEQUENCE   116 AA;  13068 MW;  5A58B9C2154D78F8 CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 116;
  Best Local Similarity   60.0%;  Pred. No. 1.6e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db     57 RTLAF 61


RESULT 9
AMCY_METEX
ID   AMCY_METEX     STANDARD;      PRT;   119 AA.
AC   P04172;
DT   20-MAR-1987 (Rel. 04, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Amicyanin-alpha precursor.
GN   MAUC.
OS   Methylobacterium extorquens.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Methylobacterium group; Methylobacterium.
OX   NCBI_TaxID=408;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AM1 / NCIMB 9133;
RX   MEDLINE=91358385; PubMed=1653226;
RA   Chistoserdov A.Y., Tsygankov Y.D., Lidstrom M.E.;
RT   "Genetic organization of methylamine utilization genes from
RT   Methylobacterium extorquens AM1.";
RL   J. Bacteriol. 173:5901-5908(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AM1 / NCIMB 9133;
RX   MEDLINE=94292425; PubMed=8021187;
RA   Chistoserdov A.Y., Chistoserdova L.V., McIntire W.S., Lidstrom M.E.;
RT   "Genetic organization of the mau gene cluster in Methylobacterium
RT   extorquens AM1: complete nucleotide sequence and generation and
RT   characteristics of mau mutants.";
RL   J. Bacteriol. 176:4052-4065(1994).
RN   [3]
RP   SEQUENCE OF 21-119.
RC   STRAIN=AM1 / NCIMB 9133;
RX   MEDLINE=86130354; PubMed=4091802;
RA   Ambler R.P., Tobari J.;
RT   "The primary structures of Pseudomonas AM1 amicyanin and
RT   pseudoazurin. Two new sequence classes of blue copper proteins.";
RL   Biochem. J. 232:451-457(1985).
CC   -!- FUNCTION: PRIMARY ACCEPTOR OF ELECTRONS FROM METHYLAMINE
CC       DEHYDROGENASE. PASSES THOSE ELECTRONS ON EITHER A SOLUBLE
CC       CYTOCHROME C OR TO PSEUDOAZURIN.
CC   -!- PATHWAY: Methylamine utilization.
CC   -!- SUBCELLULAR LOCATION: Periplasmic.
CC   -!- SIMILARITY: CONTAINS 1 PLASTOCYANIN-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M57963; AAA68895.1; -.
DR   EMBL; L26406; AAB46937.1; -.
DR   PIR; A00295; CUPSAM.
DR   HSSP; P22364; 1AAC.
DR   InterPro; IPR000923; BlueCu_1.
DR   InterPro; IPR001235; Copper_blue.
DR   Pfam; PF00127; copper-bind; 1.
DR   PRINTS; PR00156; COPPERBLUE.
DR   ProDom; PD001235; Copper_blue; 1.
DR   PROSITE; PS00196; COPPER_BLUE; 1.
KW   Copper; Electron transport; Periplasmic; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21    119       AMICYANIN-ALPHA.
FT   DOMAIN       21    119       PLASTOCYANIN-LIKE.
FT   METAL        67     67       COPPER (BY SIMILARITY).
FT   METAL       106    106       COPPER (BY SIMILARITY).
FT   METAL       109    109       COPPER (BY SIMILARITY).
FT   METAL       112    112       COPPER (BY SIMILARITY).
SQ   SEQUENCE   119 AA;  12609 MW;  732FDECA8239D857 CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 119;
  Best Local Similarity   60.0%;  Pred. No. 1.7e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db      2 RALAF 6


RESULT 10
HV2E_HUMAN
ID   HV2E_HUMAN     STANDARD;      PRT;   121 AA.
AC   P01818;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig heavy chain V-II region HE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=70114712; PubMed=5264153;
RA   Cunningham B.A., Pflumm M.N., Rutishauser U., Edelman G.M.;
RT   "Subgroups of amino acid sequences in the variable regions of
RT   immunoglobulin heavy chains.";
RL   Proc. Natl. Acad. Sci. U.S.A. 64:997-1003(1969).
CC   -!- MISCELLANEOUS: THIS GAMMA-1 CHAIN WAS ISOLATED FROM A MYELOMA
CC       PROTEIN.
DR   PIR; A02093; G1HUHE.
DR   HSSP; P01825; 7FAB.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_v.
```

```
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
KW   Immunoglobulin V region.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   NON_TER     121    121
SQ   SEQUENCE   121 AA;  13483 MW;  88A5082C273753B4 CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 121;
  Best Local Similarity   60.0%;  Pred. No. 1.7e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db     104 RTLAF 108


RESULT 11
NRDJ_BACSU
ID   NRDJ_BACSU     STANDARD;      PRT;   129 AA.
AC   O31876; O64172;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phage-derived nrdI protein (BnrdI).
GN   NRDIB.
OS   Bacillus subtilis, and
OS   Bacteriophage SPBc2.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423, 66797;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=B.subtilis; STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Phage SPBc2;
RA   Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Mauel C.,
RA   Karamata D.;
RT   "The complete nucleotide sequence of the Bacillus subtilis SPbetac2
RT   prophage.";
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: NOT KNOWN; PROBABLY INVOLVED IN RIBONUCLEOTIDE REDUCTASE
CC       FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE NRDI FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z99114; CAB13899.1; -.
DR   EMBL; AF020713; AAC13133.1; -.
DR   SubtiList; BG13722; nrdIB.
DR   InterPro; IPR004465; NrdI.
DR   TIGRFAMs; TIGR00333; nrdI; 1.
KW   Complete proteome.
SQ   SEQUENCE   129 AA;  14655 MW;  29A46D4D04613EB4 CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 129;
  Best Local Similarity   60.0%;  Pred. No. 1.8e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db      53 RTLSF 57


RESULT 12
LECA_ARTIN
ID   LECA_ARTIN     STANDARD;      PRT;   133 AA.
AC   P18670; P80023;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Agglutinin alpha chain (Jacalin alpha chain).
OS   Artocarpus integer (Jack fruit) (Artocarpus integrifolia).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Rosales; Moraceae; Artocarpus.
OX   NCBI_TaxID=3490;
RN   [1]
RP   SEQUENCE, CARBOHYDRATE-LINKAGE SITES, AND SUBUNITS.
RC   TISSUE=Seed;
RX   MEDLINE=92392266; PubMed=1520261;
RA   Ruffet E., Paquet N., Frutiger S., Hughes G.J., Jaton J.-C.;
RT   "Structural and electron-microscopic studies of jacalin from
RT   jackfruit (Artocarpus integrifolia) show that this lectin is a 65 kDa
RT   tetramer.";
RL   Biochem. J. 286:131-134(1992).
RN   [2]
RP   SEQUENCE.
RC   TISSUE=Seed;
RX   MEDLINE=91243835; PubMed=2037053;
RA   Young N.M., Johnston R.A.Z., Watson D.C.;
RT   "The amino acid sequences of jacalin and the Maclura pomifera
RT   agglutinin.";
RL   FEBS Lett. 282:382-384(1991).
RN   [3]
RP   SEQUENCE.
RC   TISSUE=Fruit;
RX   MEDLINE=92287028; PubMed=1599414;
RA   Mahanta S.K., Sanker S., Prasad Rao N.V.S.A.V., Swamy M.J.,
RA   Surolia A.;
RT   "Primary structure of a Thomsen-Friedenreich-antigen-specific lectin,
RT   jacalin [Artocarpus integrifolia (jack fruit) agglutinin]. Evidence
RT   for the presence of an internal repeat.";
RL   Biochem. J. 284:95-101(1992).
RN   [4]
RP   SEQUENCE OF 1-33.
RC   TISSUE=Seed;
RX   MEDLINE=89206218; PubMed=2705782;
RA   Young N.M., Johnston R.A.Z., Szabo A.G., Watson D.C.;
```

```
RT   "Homology of the D-galactose-specific lectins from Artocarpus
RT   integrifolia and Maclura pomifera and the role of an unusual small
RT   polypeptide subunit.";
RL   Arch. Biochem. Biophys. 270:596-603(1989).
RN   [5]
RP   SEQUENCE OF 1-29 AND 68-89.
RC   TISSUE=Seed;
RX   MEDLINE=93160237; PubMed=8431469;
RA   Kabir S., Aebersold R., Daar A.S.;
RT   "Identification of a novel 4 kDa immunoglobulin-A-binding peptide
RT   obtained by the limited proteolysis of jacalin.";
RL   Biochim. Biophys. Acta 1161:194-200(1993).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.43 ANGSTROMS).
RX   MEDLINE=96266349; PubMed=8673603;
RA   Sakaranarayanan R., Sekar S., Banerjee R., Sharma V., Surolia A.,
RA   Vijayan M.;
RT   "A novel mode of carbohydrate recognition in jacalin, a Moraceae
RT   plant lectin with a beta-prism fold.";
RL   Nat. Struct. Biol. 3:596-603(1996).
CC   -!- FUNCTION: D-GALACTOSE-SPECIFIC LECTIN, BINDS THE T-ANTIGEN
CC       STRUCTURE GAL-BETA1-3-GALNAC (THOMSEN-FRIEDENREICH-ANTIGEN-
CC       SPECIFIC LECTIN).
CC   -!- FUNCTION: POTENT AND SELECTIVE STIMULANT OF DISTINCT T- AND B-CELL
CC       FUNCTIONS. SHOWS A UNIQUE ABILITY TO SPECIFICALLY RECOGNIZE IGA-1
CC       FROM HUMAN SERUM.
CC   -!- SUBUNIT: TETRAMER OF FOUR ALPHA CHAIN ASSOCIATED WITH TWO OR FOUR
CC       BETA CHAINS.
CC   -!- SIMILARITY: TO THE MACLURA POMIFERA AGGLUTININ ALPHA CHAIN.
DR   PIR; S03989; S03989.
DR   PIR; S15824; S15824.
DR   PIR; S21291; S21291.
DR   PIR; S24429; S24429.
DR   PDB; 1JAC; 05-JUN-97.
DR   InterPro; IPR001229; Jacalin_lectin.
DR   Pfam; PF01419; Jacalin; 1.
KW   Lectin; Glycoprotein; Repeat; IgA-binding protein; 3D-structure.
FT   REPEAT        7     64
FT   REPEAT       76    130
FT   DOMAIN       68     89       IGA-BINDING.
FT   CARBOHYD     35     35       N-LINKED (GLCNAC. . .) (PARTIAL).
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .); IN ONE FORM.
FT   VARIANT      31     31       V -> I.
FT   VARIANT      34     34       L -> G.
FT   VARIANT      45     45       K -> L OR T.
FT   VARIANT      66     66       M -> D OR V.
FT   VARIANT      67     67       E -> M.
FT   VARIANT      72     72       T -> I.
FT   VARIANT      74     74       N -> K.
FT   VARIANT     102    102       T -> D.
FT   VARIANT     113    113       I -> V.
FT   VARIANT     131    131       L -> N.
FT   CONFLICT     75     75       V -> Y (IN REF. 3).
SQ   SEQUENCE   133 AA;  14662 MW;  FF10513379CB2E10 CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 133;
  Best Local Similarity   60.0%;  Pred. No. 1.9e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db     82 RSLTF 86


RESULT 13
LECA_MACPO
ID   LECA_MACPO     STANDARD;      PRT;   133 AA.
AC   P18674;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin alpha chain (MPA).
OS   Maclura pomifera (Osage orange).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Rosales; Moraceae; Maclura.
OX   NCBI_TaxID=3496;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Seed;
RX   MEDLINE=91243835; PubMed=2037053;
RA   Young N.M., Johnston R.A.Z., Watson D.C.;
RT   "The amino acid sequences of jacalin and the Maclura pomifera
RT   agglutinin.";
RL   FEBS Lett. 282:382-384(1991).
RN   [2]
RP   SEQUENCE OF 1-33.
RC   TISSUE=Seed;
RX   MEDLINE=89206218; PubMed=2705782;
RA   Young N.M., Johnston R.A.Z., Szabo A.G., Watson D.C.;
RT   "Homology of the D-galactose-specific lectins from Artocarpus
RT   integrifolia and Maclura pomifera and the role of an unusual small
RT   polypeptide subunit.";
RL   Arch. Biochem. Biophys. 270:596-603(1989).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX   MEDLINE=98165814; PubMed=9497359;
RA   Lee X., Thompson A., Zhang Z., Ton-That H., Biesterfeldt J., Ogata C.,
RA   Xu L., Johnston R.A., Young N.M.;
RT   "Structure of the complex of Maclura pomifera agglutinin and the T-
RT   antigen disaccharide, Galbeta1,3GalNAc.";
RL   J. Biol. Chem. 273:6312-6318(1998).
CC   -!- FUNCTION: D-GALACTOSE-SPECIFIC LECTIN, BINDS THE T-ANTIGEN
CC       STRUCTURE GAL-BETA1-3-GALNAC.
CC   -!- SUBUNIT: FORMED OF FOUR ALPHA CHAINS AND FOUR BETA CHAINS.
CC   -!- SIMILARITY: TO THE ARTOCARPUS INTEGER AGGLUTININ ALPHA CHAIN.
DR   PIR; S03990; S03990.
DR   PIR; S15825; S15825.
DR   PDB; 1JOT; 16-FEB-99.
DR   InterPro; IPR001229; Jacalin_lectin.
DR   Pfam; PF01419; Jacalin; 1.
KW   Lectin; 3D-structure.
FT   VARIANT      31     31       T -> V (IN MINOR FORMS).
FT   VARIANT      52     52       K -> T (IN MINOR FORMS).
FT   VARIANT      59     59       E -> D (IN MINOR FORMS).
FT   VARIANT      72     72       V -> I (IN MINOR FORMS).
FT   VARIANT      81     81       I -> V (IN MINOR FORMS).
FT   VARIANT     110    110       N -> Q (IN MINOR FORMS).
FT   VARIANT     112    112       L -> G (IN MINOR FORMS).
FT   CONFLICT     29     29       R -> D (IN REF. 2).
FT   CONFLICT     32     33       YD -> QN (IN REF. 2).
SQ   SEQUENCE   133 AA;  14758 MW;  15C69FF94B6D09FD CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 133;
  Best Local Similarity   60.0%;  Pred. No. 1.9e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 RXLXF 8
          | | |
Db     82 RSLTF 86


RESULT 14
V223_FOWPV
ID   V223_FOWPV     STANDARD;      PRT;   141 AA.
AC   Q9J512;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative ankyrin-repeat protein FPV223.
GN   FPV223.
OS   Fowlpox virus (FPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Avipoxvirus.
OX   NCBI_TaxID=10261;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20193820; PubMed=10729156;
RA   Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT   "The genome of fowlpox virus.";
RL   J. Virol. 74:3815-3831(2000).
CC   -!- SIMILARITY: CONTAINS 4 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF198100; AAF44567.1; -.
DR   HSSP; Q13625; 1YCS.
DR   InterPro; IPR002110; ANK.
DR   Pfam; PF00023; ank; 5.
DR   SMART; SM00248; ANK; 2.
DR   PROSITE; PS50088; ANK_REPEAT; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   Hypothetical protein; Repeat; ANK repeat.
FT   REPEAT       21     50       ANK 1.
FT   REPEAT       54     83       ANK 2.
FT   REPEAT       85    114       ANK 3.
FT   REPEAT      118    140       ANK 4.
SQ   SEQUENCE   141 AA;  16075 MW;  D29ED1CFD067C12E CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 141;
  Best Local Similarity   60.0%;  Pred. No. 2e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db      36 RALSF 40


RESULT 15
YEF5_YEAST
ID   YEF5_YEAST     STANDARD;      PRT;   141 AA.
AC   P32616;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   Hypothetical 16.5 kDa protein in GLY1-GDA1 intergenic region.
GN   YEL045C OR SYGP-ORF33.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mulligan J.T., Dietrich F.S., Hennessey K.M., Sehl P., Komp C.,
RA   Wei Y., Taylor P., Nakahara K., Roberts D., Davis R.W.;
RL   Submitted (FEB-1993) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA   Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA   Chung E., Duncan M., Guzman E., Hartzell G., Hunicke-Smith S.,
RA   Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA   Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA   Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA   Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U18779; AAB64997.1; -.
DR   PIR; S30832; S30832.
DR   SGD; S0000771; YEL045C.
KW   Hypothetical protein; ATP-binding; Transmembrane.
FT   NP_BIND      15     22       ATP (POTENTIAL).
FT   TRANSMEM     38     58       POTENTIAL.
FT   TRANSMEM     67     87       POTENTIAL.
SQ   SEQUENCE   141 AA;  16468 MW;  F6604AC5343A5D5C CRC64;

  Query Match             75.0%;  Score 15;  DB 1;  Length 141;
  Best Local Similarity   60.0%;  Pred. No. 2e+02;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 RXLXF 8
           | | |
Db       7 RTLAF 11


Search completed: December 14, 2002, 15:46:43
Job time : 26.5 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: December 14, 2002, 15:45:49 ; Search time 30.5 Seconds
(without alignments)
25.216 Million cell updates/sec

Title: US-09-726-470A-2
Perfect score: 20
Sequence: 1 XXXRXLXF 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	15	75.0	22	2	S47206	T-cell receptor J-
2	15	75.0	23	2	S47192	T-cell receptor J-
3	15	75.0	39	2	B85990	hypothetical prote
4	15	75.0	42	2	T07581	hypothetical prote
5	15	75.0	57	2	S16587	hypothetical prote
6	15	75.0	60	2	D90971	hypothetical prote
7	15	75.0	60	2	D85744	unknown protein en
8	15	75.0	63	2	AG1332	hypothetical prote
9	15	75.0	63	2	AG1703	hypothetical prote
10	15	75.0	77	2	JC5052	hypothetical 8.9k
11	15	75.0	79	2	B83400	hypothetical prote
12	15	75.0	79	2	I38991	tumorigenic conver
13	15	75.0	80	2	T39148	hypothetical prote
14	15	75.0	80	2	B43259	H+-transporting tw
15	15	75.0	86	2	S35769	T-cell receptor al
16	15	75.0	93	2	AC0592	probable membrane
17	15	75.0	97	2	B90725	hypothetical prote
18	15	75.0	97	2	C85576	hypothetical prote
19	15	75.0	97	2	F64809	ybgE protein - Esc
20	15	75.0	98	2	A95329	probable fragment
21	15	75.0	100	2	E69846	hypothetical prote
22	15	75.0	101	2	S51384	hypothetical prote
23	15	75.0	101	2	AE1386	transcription regu
24	15	75.0	101	2	AG1761	transcription regu
25	15	75.0	104	2	A70665	probable ureB prot
26	15	75.0	109	2	G64609	hypothetical prote
27	15	75.0	110	2	C55228	hypothetical prote
28	15	75.0	113	2	S55528	Ig heavy chain V r
29	15	75.0	113	2	S55530	Ig heavy chain V r
30	15	75.0	113	2	S55533	Ig heavy chain V r
31	15	75.0	113	2	S55531	Ig heavy chain V r
32	15	75.0	113	2	S55532	Ig heavy chain V r
33	15	75.0	113	2	F72687	hypothetical prote
34	15	75.0	114	2	D71048	hypothetical prote
35	15	75.0	116	2	G82537	pterin-4-alpha-car
36	15	75.0	119	1	CUPSAM	amicyanin precurso
37	15	75.0	119	2	E72714	probable ribosomal
38	15	75.0	121	1	G1HUHE	Ig heavy chain V-I
39	15	75.0	122	2	H82231	hypothetical prote
40	15	75.0	128	2	E70547	hypothetical prote
41	15	75.0	129	2	T12924	conserved hypothet
42	15	75.0	129	2	H72627	hypothetical prote
43	15	75.0	133	2	D48776	polyprotein (E2/NS
44	15	75.0	133	2	B30242	stem cell protein
45	15	75.0	133	2	S15825	agglutinin alpha c

ALIGNMENTS

RESULT 1
S47206
T-cell receptor J-alpha wnVII.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 23-Jul-1999
C;Accession: S47206
R;Plaza, A.; Kono, D.H.; Theofilopoulos, A.N.
submitted to the EMBL Data Library, February 1993
A;Reference number: S40133
A;Accession: S47206
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-22 <PLA>
A;Cross-references: EMBL:X71036; NID:g507043; PIDN:CAA50353.1; PID:g510651
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: T-cell receptor

Query Match 75.0%; Score 15; DB 2; Length 22;
Best Local Similarity 60.0%; Pred. No. 83;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 8 RALTF 12

RESULT 2
S47192
T-cell receptor J-alpha wnVII.2 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 23-Jul-1999
C;Accession: S47192
R;Plaza, A.; Kono, D.H.; Theofilopoulos, A.N.
submitted to the EMBL Data Library, February 1993
A;Reference number: S40133
A;Accession: S47192
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-23 <PLA>
A;Cross-references: EMBL:X71051; NID:g506974; PIDN:CAA50368.1; PID:g510653
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: T-cell receptor

Query Match 75.0%; Score 15; DB 2; Length 23;
Best Local Similarity 60.0%; Pred. No. 87;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 9 RALTF 13

RESULT 3
B85990
hypothetical protein Z4614 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: B85990
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: B85990
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-39 <STO>
A;Cross-references: GB:AE005174; NID:g12517881; PIDN:AAG58382.1; GSPDB:GN00145; UWGP:Z46
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z4614

Query Match 75.0%; Score 15; DB 2; Length 39;
Best Local Similarity 60.0%; Pred. No. 1.5e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 28 RALAF 32

RESULT 4
T07581
hypothetical protein 42g - Japanese black pine chloroplast
C;Species: chloroplast Pinus thunbergiana (Japanese black pine)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 18-Aug-2000
C;Accession: T07581
R;Wakasugi, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Tsudzuki, T.; Sugiura, M.
Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994
A;Title: Loss of all ndh genes as determined by sequencing the entire chloroplast genome
A;Reference number: Z16030; MUID:95024047; PMID:7937893
A;Accession: T07581
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-42 <WAK>
A;Cross-references: EMBL:D17510; NID:g529643; PIDN:BAA04457.1; PID:g1262742
C;Genetics:
A;Genome: chloroplast
C;Keywords: chloroplast

Query Match 75.0%; Score 15; DB 2; Length 42;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 14 RSLSF 18

RESULT 5
S16587
hypothetical protein 1 - lamb's-quarters
C;Species: Chenopodium album (lamb's-quarters)
C;Date: 21-Nov-1993 #sequence_revision 26-May-1995 #text_change 26-May-1995
C;Accession: S16587
R;Doerfel, P.; Weihe, A.; Dolferus, R.; Boerner, T.
Plant Mol. Biol. 17, 155-156, 1991
A;Title: DNA sequence of a mitochondrial plasmid from Chenopodium album.
A;Reference number: S16587; MUID:91329724; PMID:1651127
A;Accession: S16587
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-57 <DOE>
A;Cross-references: EMBL:X58911

Query Match 75.0%; Score 15; DB 2; Length 57;
Best Local Similarity 60.0%; Pred. No. 2.1e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 26 RTLTF 30

RESULT 6
D90971
hypothetical protein ECs2740 [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: D90971
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90971
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-60 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB36163.1; PID:g13362208; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs2740

Query Match 75.0%; Score 15; DB 2; Length 60;
Best Local Similarity 60.0%; Pred. No. 2.2e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 29 RALAF 33

RESULT 7
D85744
unknown protein encoded within prophage CP-933R [imported] - Escherichia coli (strain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: D85744
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85744
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-60 <STO>
A;Cross-references: GB:AE005174; NID:g12515365; PIDN:AAG56416.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z2370

Query Match 75.0%; Score 15; DB 2; Length 60;
Best Local Similarity 60.0%; Pred. No. 2.2e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 29 RALAF 33

RESULT 8
AG1332
hypothetical protein lmo2063 [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AG1332

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1332
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-63 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAD00141.1; PID:g16411533; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo2063

  Query Match 75.0%; Score 15; DB 2; Length 63;
  Best Local Similarity 60.0%; Pred. No. 2.3e+02;
  Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 39 RTLTF 43

RESULT 9
AG1703
hypothetical protein lin2169 [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AG1703
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1703
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-63 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC97399.1; PID:g16414683; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin2169

  Query Match 75.0%; Score 15; DB 2; Length 63;
  Best Local Similarity 60.0%; Pred. No. 2.3e+02;
  Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 39 RTLTF 43

RESULT 10
JC5052
hypothetical 8.9k protein - Shigella flexneri
C;Species: Shigella flexneri
C;Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 26-Aug-1999
C;Accession: JC5052
R;Venkatesan, M.M.; Alexander, W.A.; Fernandez-Prada, C.
Gene 175, 23-27, 1996
A;Title: A Shigella flexneri invasion plasmid gene, ipgH, with homology to IS629 and seq
A;Reference number: JC5050; MUID:97074644; PMID:8917071
A;Accession: JC5052
A;Molecule type: DNA
A;Residues: 1-77 <VEN>
A;Cross-references: GB:U28354; NID:g1016674; PIDN:AAC44576.1; PID:g1016677
A;Note: in the authors' translation, residues 5-7 are shown after residue 15, residues 8
C;Superfamily: DNA replication protein dnaC

  Query Match 75.0%; Score 15; DB 2; Length 77;
  Best Local Similarity 60.0%; Pred. No. 2.8e+02;
  Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 49 RSLSF 53

RESULT 11
B83400
hypothetical protein PA1970 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: B83400
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: B83400
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-79 <STO>
A;Cross-references: GB:AE004623; GB:AE004091; NID:g9947961; PIDN:AAG05358.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1970

  Query Match 75.0%; Score 15; DB 2; Length 79;
  Best Local Similarity 60.0%; Pred. No. 2.9e+02;
  Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 44 RSLSF 48

RESULT 12
I38991
tumorigenic conversion-associated protein CATR1 - human
C;Species: Homo sapiens (man)
C;Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 29-Aug-1997
C;Accession: I38991
R;Li, D.; Noyes, I.; Shuler, C.; Milo, G.E.
Proc. Natl. Acad. Sci. U.S.A. 92, 6409-6413, 1995
A;Title: Cloning and sequencing of CATR1.3, a human gene associated with tumorigenic
A;Reference number: I38991; MUID:95327656; PMID:7604004
A;Accession: I38991
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-79 <RES>
A;Cross-references: EMBL:U25433; NID:g896044; PID:g896045
C;Genetics:
A;Gene: GDB:CATR1
A;Cross-references: GDB:633071; OMIM:600676
A;Map position: 16p13.3-16p13.3

  Query Match 75.0%; Score 15; DB 2; Length 79;
  Best Local Similarity 60.0%; Pred. No. 2.9e+02;
  Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 42 RALTF 46

RESULT 13
T39148
hypothetical protein SPAC8C9.11 - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T39148
R;Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21748
A;Accession: T39148
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-80 <OLI>
A;Cross-references: EMBL:Z99168; PIDN:CAB16299.1; GSPDB:GN00066; SPDB:SPAC8C9.11
A;Experimental source: strain 972h-; cosmid c8C9
C;Genetics:
A;Gene: SPDB:SPAC8C9.11
A;Map position: 1
A;Introns: 20/3

  Query Match 75.0%; Score 15; DB 2; Length 80;
  Best Local Similarity 60.0%; Pred. No. 2.9e+02;
  Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
     | | |
Db 72 RTLSF 76

RESULT 14
B43259
H+-transporting two-sector ATPase (EC 3.6.3.14) chain a - Enterococcus hirae (fragment)
C;Species: Enterococcus hirae
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 03-Jun-2002
C;Accession: B43259
R;Shibata, C.; Ehara, T.; Tomura, K.; Igarashi, K.; Kobayashi, H.
J. Bacteriol. 174, 6117-6124, 1992
A;Title: Gene structure of Enterococcus hirae (Streptococcus faecalis) F1F0-ATPase, whic
A;Reference number: A43259; MUID:93015650; PMID:1328152
A;Accession: B43259
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-80 <SHI>
A;Experimental source: ATCC 9790
A;Note: sequence extracted from NCBI backbone (NCBIN:115116, NCBIP:115124)
C;Keywords: hydrolase

  Query Match 75.0%; Score 15; DB 2; Length 80;
  Best Local Similarity 60.0%; Pred. No. 2.9e+02;
  Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
     | | |
Db 4 RSLTF 8

RESULT 15
S35769
T-cell receptor alpha chain - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 23-Jul-1999
C;Accession: S35769
R;Wedderburn, L.R.
submitted to the EMBL Data Library, June 1993
A;Reference number: S35769
A;Accession: S35769
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-86 <WED>
A;Cross-references: EMBL:Z22965; NID:g312153; PIDN:CAA80538.1; PID:g312154
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: T-cell receptor

  Query Match 75.0%; Score 15; DB 2; Length 86;
  Best Local Similarity 60.0%; Pred. No. 3.2e+02;
  Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
     | | |
Db 56 RALTF 60

Search completed: December 14, 2002, 15:50:06
Job time : 33.5 secs

GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2002, 15:50:04 ; Search time 220 Seconds
(without alignments)
81.891 Million cell updates/sec

Title: US-09-726-470A-2
Perfect score: 20
Sequence: 1 XXXRXLXF 8

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09726470/runat_10122002_090717_4962/app_query.fasta_1.398
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09726470_@CGN_1_1_79_@runat_10122002_090717_4962 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		15	75.0	15	17	AAT41973	HIV-1 gag binding
2		15	75.0	15	18	AAT94864	HIV-1 gag gene ant
3		15	75.0	16	17	AAT41974	HIV-1 gag binding
4		15	75.0	16	18	AAT94865	HIV-1 gag gene ant
5		15	75.0	17	17	AAT41975	HIV-1 gag binding
6		15	75.0	17	18	AAT94867	HIV-1 gag gene ant
7	c	15	75.0	18	18	AAT47720	Mouse bone morphog
8	c	15	75.0	18	21	AAZ44143	Human EGR-1 DNA an
9		15	75.0	19	17	AAT41998	HIV-1 gag binding
10		15	75.0	19	18	AAT94890	HIV-1 gag gene ant
11		15	75.0	20	19	AAV40304	Maize oligonucleot
12		15	75.0	20	20	AAZ05921	PCR primer used to
13		15	75.0	20	20	AAZ05924	PCR primer used to
14		15	75.0	20	20	AAX97011	PCR primer used to
15	c	15	75.0	20	20	AAX94959	PCR primer used to
16		15	75.0	20	21	AAZ70026	Human biallelic ma
17	c	15	75.0	20	21	AAA96274	Sequence of a stab
18		15	75.0	20	21	AAA96280	Sequence of a stab
19		15	75.0	20	21	AAC79561	Human p38alpha ant
20	c	15	75.0	20	21	AAA59794	Primer for TRIP 1
21		15	75.0	20	21	AAA40879	Murine TNFalpha an
22		15	75.0	20	21	AAA36753	Human dysferlin re
23		15	75.0	20	21	AAA36771	Human dysferlin re
24		15	75.0	20	21	AAX82844	Human dysferlin PC
25		15	75.0	20	21	AAX82862	Human dysferlin PC
26		15	75.0	20	21	AAZ39096	Human mcl-1 anti-a
27		15	75.0	20	22	AAH27931	PCR primer for a m
28	c	15	75.0	20	22	AAH41776	TRIP 1 gene PCR pr
29	c	15	75.0	20	24	AAS62929	Esophageal adenoca
30		15	75.0	21	15	AAQ74796	Primer for amplify
31	c	15	75.0	21	18	AAT51776	Fibrillin 1 Fbn1 g
32		15	75.0	21	22	AAF95403	Human gene single
33		15	75.0	21	23	ABA10032	Tail primer #25 fr
34	c	15	75.0	21	24	ABL60754	Neurofibromatosis
35	c	15	75.0	21	24	AAS97487	Murine SAC1 gene-s
36	c	15	75.0	22	19	AAV05177	Primer JP72 used t
37		15	75.0	23	16	AAQ94778	3' Oligonucleotide
38		15	75.0	23	24	AAD37030	3' PCR primer used
39	c	15	75.0	23	24	ABA92646	Thiolated oligonuc
40	c	15	75.0	23	24	ABA92647	Thiolated oligonuc
41		15	75.0	24	13	AAQ29499	EDA-FN primer (2).
42		15	75.0	24	16	AAT00087	Hepatitis GB virus
43		15	75.0	24	21	AAA55333	Hepatitis GB virus
44		15	75.0	24	21	AAZ88091	HIV gag AUG mutant
45	c	15	75.0	24	22	AAI71874	Cytochrome P45011b

ALIGNMENTS

RESULT 1
AAT41973
ID AAT41973 standard; cDNA; 15 BP.
XX
AC AAT41973;
XX
DT 24-JUN-1997 (first entry)
XX
DE HIV-1 gag binding oligonucleotide used as antisense HIV inhibitor.
XX
KW Co-operative binding; duplex; antisense inhibition; target sequence;
KW human immunodeficiency virus; HIV; dimerisation domain; T structure;
KW gene function; ss.
XX
OS Synthetic.
XX
PN WO9632474-A1.
XX

```
PD   17-OCT-1996.
XX
PF   04-APR-1996;   96WO-US04605.
XX
PR   12-APR-1995;   95US-0420672.
XX
PA   (HYBR-) HYBRIDON INC.
XX
PI   Agrawal S,  Kandimalla ER;
XX
DR   WPI; 1996-477125/47.
XX
PT   Compsn. contg. at least two co-operative oligo:nucleotide(s)
PT   complementary to a target sequence  - and with mutually
PT   complementary dimerisation domains, for use as antisense inhibitors
PT   of HIV and influenza virus
XX
PS   Disclosure; Page 44; 84pp; English.
XX
CC   AAT41965-T41994 are oligonucleotides (ON) that bind to HIV-1 mRNA and/or
CC   DNA and act as antisense inhibitors of HIV-1 gene expression. The
CC   ON are preferably used as duplexes, i.e. a first ON has a region
CC   that binds to an HIV-1 target sequence in a 5'-3' direction and a
CC   second region complementary to a second ON which has a first region
CC   which binds to the same target HIV-1 sequence but in a 3'-5' direction
CC   and a second region complementary to the first ON. Both ON bind to the
CC   HIV target sequence up until a certain point along the target sequence,
CC   where the two binding ON are in close proximity and the remainder of
CC   the binding ON bind each other. The duplex/target site complex forms a
CC   T-shaped structure, inhibiting expression of HIV-1 nucleic acid. Also
CC   co-operative ON, when labelled, can be used to identify specific
CC   bacteria or viruses in cell cultures; to study function of specific genes
CC   as an alternative to use of 'knock out' animals. Co-operative ON have
CC   improved affinity and sequence specificity, reduced toxicity and better
CC   antisense activity compared with single, longer ON. Insertion of
CC   dimerisation domains into antisense ON (i.e. sequences hybridising to a
CC   second antisense ON) provides a more stable complex.
XX
SQ   Sequence 15 BP; 0 A; 7 C; 2 G; 6 T; 0 other;

Alignment Scores:
Pred. No.:                1.01e+03        Length:         15
Score:                    15.00           Matches:        3
Percent Similarity:       60.00%          Conservative:   0
Best Local Similarity:    60.00%          Mismatches:     2
Query Match:              75.00%          Indels:         0
DB:                       17              Gaps:           0

US-09-726-470A-2 (1-8) x AAT41973 (1-15)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      1 CGGTCTCTCTCCTTC 15

RESULT 2
AAT94864
ID   AAT94864 standard; cDNA; 15 BP.
XX
AC   AAT94864;
XX
DT   22-APR-1998  (first entry)
XX
DE   HIV-1 gag gene antisense oligonucleotide.
XX
KW   gag gene; initiation codon region; target region; dimerisation domain;
KW   synthetic cooperative oligonucleotide; affinity; specificity;
KW   antisense molecule; treatment; viral infection; influenza; HIV; ss.
XX
OS   Synthetic.
OS   Human immunodeficiency virus type 1.
XX
FH   Key             Location/Qualifiers
FT   misc_binding     1..3
FT                    /*tag=  a
FT                    /note= "dimerisation domain which hybridises to
FT                           nucleotides 1-3 of AAT94869"
FT   misc_binding     4..15
FT                    /*tag=  b
FT                    /note= "binds to HIV-1 gag target AAT94853"
XX
PN   WO9738097-A1.
XX
PD   16-OCT-1997.
XX
PF   04-APR-1997;   97WO-US05683.
XX
PR   04-APR-1996;   96US-0627967.
XX
PA   (HYBR-) HYBRIDON INC.
XX
PI   Agrawal S,  Kandimalla ER;
XX
DR   WPI; 1997-512714/47.
XX
PT   Anti:sense oligo:nucleotide compositions have at least 2 cooperative
PT   oligo:nucleotide(s) having a targeting and a dimerisation region -
PT   useful for inhibition of target nucleic acid expression
XX
PS   Disclosure; Page 26; 91pp; English.
XX
CC   The present oligonucleotide is an antisense oligonucleotide that binds
CC   to part of the gag gene of Human immunodeficiency virus type 1 (HIV-1).
CC   The present oligonucleotide has an extended sequence at the 5' end of
CC   the binding sequence which forms a duplex stem with the corresponding
CC   antisense dimerisation domain of AAT94869, when the 2 oligonucleotides
CC   bind to adjacent sites on the target sequence. The stability of binding
CC   was found to increase with the number of bases in the dimerisation
CC   domain. The oligonucleotides are used to exemplify the method of the
CC   invention. This method comprises two synthetic cooperative
CC   oligonucleotides, where each oligonucleotide comprises a region
CC   complementary to one of tandem, non-overlapping regions of a target
CC   nucleic acid, and a dimerisation domain at a terminus of each of the
CC   oligonucleotides. The dimerisation domains of the oligonucleotides are
CC   complementary to each other. The target nucleic acid is an mRNA, a
CC   single-stranded viral DNA, or a single-stranded viral RNA. The synthetic
CC   oligonucleotides can interact cooperatively to provide improved
CC   affinity, specificity, and biological activity as antisense molecules.
CC   The compositions are used for inhibiting the expression of target
CC   nucleic acids. They can be used for the treatment of viral infections,
CC   e.g. influenza (AAV04801-17) or HIV (AAT94853-92) infection. They can
CC   also be used for the detection and study of target nucleic acids.
XX
SQ   Sequence 15 BP; 0 A; 7 C; 2 G; 6 T; 0 other;

Alignment Scores:
Pred. No.:                1.01e+03        Length:         15
Score:                    15.00           Matches:        3
Percent Similarity:       60.00%          Conservative:   0
Best Local Similarity:    60.00%          Mismatches:     2
Query Match:              75.00%          Indels:         0
DB:                       18              Gaps:           0

US-09-726-470A-2 (1-8) x AAT94864 (1-15)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      1 CGGTCTCTCTCCTTC 15

RESULT 3
AAT41974
ID   AAT41974 standard; cDNA; 16 BP.
XX
AC   AAT41974;
XX
DT   24-JUN-1997  (first entry)
XX
```

```
DE   HIV-1 gag binding oligonucleotide used as antisense HIV inhibitor.
XX
KW   Co-operative binding; duplex; antisense inhibition; target sequence;
KW   human immunodeficiency virus; HIV; dimerisation domain; T structure;
KW   gene function; ss.
XX
OS   Synthetic.
XX
PN   WO9632474-A1.
XX
PD   17-OCT-1996.
XX
PF   04-APR-1996;   96WO-US04605.
XX
PR   12-APR-1995;   95US-0420672.
XX
PA   (HYBR-) HYBRIDON INC.
XX
PI   Agrawal S,  Kandimalla ER;
XX
DR   WPI; 1996-477125/47.
XX
PT   Compsn. contg. at least two co-operative oligo:nucleotide(s)
PT   complementary to a target sequence  - and with mutually
PT   complementary dimerisation domains, for use as antisense inhibitors
PT   of HIV and influenza virus
XX
PS   Disclosure; Page 45; 84pp; English.
XX
CC   AAT41965-T41994 are oligonucleotides (ON) that bind to HIV-1 mRNA and/or
CC   DNA and act as antisense inhibitors of HIV-1 gene expression. The
CC   ON are preferably used as duplexes, i.e. a first ON has a region
CC   that binds to an HIV-1 target sequence in a 5'-3' direction and a
CC   second region complementary to a second ON which has a first region
CC   which binds to the same target HIV-1 sequence but in a 3'-5' direction
CC   and a second region complementary to the first ON. Both ON bind to the
CC   HIV target sequence up until a certain point along the target sequence,
CC   where the two binding ON are in close proximity and the remainder of
CC   the binding ON bind each other. The duplex/target site complex forms a
CC   T-shaped structure, inhibiting expression of HIV-1 nucleic acid. Also
CC   co-operative ON, when labelled, can be used to identify specific
CC   bacteria or viruses in cell cultures; to study function of specific genes
CC   as an alternative to use of 'knock out' animals. Co-operative ON have
CC   improved affinity and sequence specificity, reduced toxicity and better
CC   antisense activity compared with single, longer ON. Insertion of
CC   dimerisation domains into antisense ON (i.e. sequences hybridising to a
CC   second antisense ON) provides a more stable complex.
XX
SQ   Sequence 16 BP; 0 A; 8 C; 2 G; 6 T; 0 other;

Alignment Scores:
Pred. No.:              1.07e+03        Length:         16
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     17              Gaps:           0

US-09-726-470A-2 (1-8) x AAT41974 (1-16)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      2 CGGTCTCTCTCCTTC 16

RESULT 4
AAT94865
ID   AAT94865 standard; cDNA; 16 BP.
XX
AC   AAT94865;
XX
DT   22-APR-1998  (first entry)
XX
DE   HIV-1 gag gene antisense oligonucleotide.
XX
KW   gag gene; initiation codon region; target region; dimerisation domain;
KW   synthetic cooperative oligonucleotide; affinity; specificity;
KW   antisense molecule; treatment; viral infection; influenza; HIV; ss.
XX
OS   Synthetic.
OS   Human immunodeficiency virus type 1.
XX
FH   Key             Location/Qualifiers
FT   misc_binding    1..4
FT                   /*tag=  a
FT                   /note= "dimerisation domain which hybridises to
FT                           nucleotides 10-13 of AAT94870"
FT   misc_binding    5..16
FT                   /*tag=  b
FT                   /note= "binds to HIV-1 gag target AAT94853"
XX
PN   WO9738097-A1.
XX
PD   16-OCT-1997.
XX
PF   04-APR-1997;   97WO-US05683.
XX
PR   04-APR-1996;   96US-0627967.
XX
PA   (HYBR-) HYBRIDON INC.
XX
PI   Agrawal S,  Kandimalla ER;
XX
DR   WPI; 1997-512714/47.
XX
PT   Anti:sense oligo:nucleotide compositions have at least 2 cooperative
PT   oligo:nucleotide(s) having a targeting and a dimerisation region -
PT   useful for inhibition of target nucleic acid expression
XX
PS   Disclosure; Page 26; 91pp; English.
XX
CC   The present oligonucleotide is an antisense oligonucleotide that binds
CC   to part of the gag gene of Human immunodeficiency virus type 1 (HIV-1).
CC   The present oligonucleotide has an extended sequence at the 5' end of
CC   the binding sequence which forms a duplex stem with the corresponding
CC   antisense dimerisation domain of AAT94870, when the 2 oligonucleotides
CC   bind to adjacent sites on the target sequence. The stability of binding
CC   was found to increase with the number of bases in the dimerisation
CC   domain. The oligonucleotides are used to exemplify the method of the
CC   invention. This method comprises two synthetic cooperative
CC   oligonucleotides, where each oligonucleotide comprises a region
CC   complementary to one of tandem, non-overlapping regions of a target
CC   nucleic acid, and a dimerisation domain at a terminus of each of the
CC   oligonucleotides. The dimerisation domains of the oligonucleotides are
CC   complementary to each other. The target nucleic acid is an mRNA, a
CC   single-stranded viral DNA, or a single-stranded viral RNA. The synthetic
CC   oligonucleotides can interact cooperatively to provide improved
CC   affinity, specificity, and biological activity as antisense molecules.
CC   The compositions are used for inhibiting the expression of target
CC   nucleic acids. They can be used for the treatment of viral infections,
CC   e.g. influenza (AAV04801-17) or HIV (AAT94853-92) infection. They can
CC   also be used for the detection and study of target nucleic acids.
XX
SQ   Sequence 16 BP; 0 A; 8 C; 2 G; 6 T; 0 other;

Alignment Scores:
Pred. No.:              1.07e+03        Length:         16
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     18              Gaps:           0

US-09-726-470A-2 (1-8) x AAT94865 (1-16)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
```

```
Db      2 CGGTCTCTCTCCTTC 16

RESULT 5
AAT41975
ID   AAT41975 standard; cDNA; 17 BP.
XX
AC   AAT41975;
XX
DT   24-JUN-1997  (first entry)
XX
DE   HIV-1 gag binding oligonucleotide used as antisense HIV inhibitor.
XX
KW   Co-operative binding; duplex; antisense inhibition; target sequence;
KW   human immunodeficiency virus; HIV; dimerisation domain; T structure;
KW   gene function; ss.
XX
OS   Synthetic.
XX
PN   WO9632474-A1.
XX
PD   17-OCT-1996.
XX
PF   04-APR-1996;   96WO-US04605.
XX
PR   12-APR-1995;   95US-0420672.
XX
PA   (HYBR-) HYBRIDON INC.
XX
PI   Agrawal S,  Kandimalla ER;
XX
DR   WPI; 1996-477125/47.
XX
PT   Compsn. contg. at least two co-operative oligo:nucleotide(s)
PT   complementary to a target sequence  - and with mutually
PT   complementary dimerisation domains, for use as antisense inhibitors
PT   of HIV and influenza virus
XX
PS   Disclosure; Page 45; 84pp; English.
XX
CC   AAT41965-T41994 are oligonucleotides (ON) that bind to HIV-1 mRNA and/or
CC   DNA and act as antisense inhibitors of HIV-1 gene expression. The
CC   ON are preferably used as duplexes, i.e. a first ON has a region
CC   that binds to an HIV-1 target sequence in a 5'-3' direction and a
CC   second region complementary to a second ON which has a first region
CC   which binds to the same target HIV-1 sequence but in a 3'-5' direction
CC   and a second region complementary to the first ON. Both ON bind to the
CC   HIV target sequence up until a certain point along the target sequence,
CC   where the two binding ON are in close proximity and the remainder of
CC   the binding ON bind each other. The duplex/target site complex forms a
CC   T-shaped structure, inhibiting expression of HIV-1 nucleic acid. Also
CC   co-operative ON, when labelled, can be used to identify specific
CC   bacteria or viruses in cell cultures; to study function of specific genes
CC   as an alternative to use of 'knock out' animals. Co-operative ON have
CC   improved affinity and sequence specificity, reduced toxicity and better
CC   antisense activity compared with single, longer ON. Insertion of
CC   dimerisation domains into antisense ON (i.e. sequences hybridising to a
CC   second antisense ON) provides a more stable complex.
XX
SQ   Sequence 17 BP; 0 A; 8 C; 3 G; 6 T; 0 other;

Alignment Scores:
Pred. No.:              1.14e+03        Length:         17
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     17              Gaps:           0

US-09-726-470A-2 (1-8) x AAT41975 (1-17)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      3 CGGTCTCTCTCCTTC 17
```

```
RESULT 6
AAT94867
ID   AAT94867 standard; cDNA; 17 BP.
XX
AC   AAT94867;
XX
DT   22-APR-1998  (first entry)
XX
DE   HIV-1 gag gene antisense oligonucleotide.
XX
KW   gag gene; initiation codon region; target region; dimerisation domain;
KW   synthetic cooperative oligonucleotide; affinity; specificity;
KW   antisense molecule; treatment; viral infection; influenza; HIV; ss.
XX
OS   Synthetic.
OS   Human immunodeficiency virus type 1.
XX
FH   Key             Location/Qualifiers
FT   misc_binding    1..5
FT                   /*tag=  a
FT                   /note= "dimerisation domain which hybridises to
FT                          nucleotides 10-14 of AAT94871"
FT   misc_binding    6..17
FT                   /*tag=  b
FT                   /note= "binds to HIV-1 gag target AAT94853"
XX
PN   WO9738097-A1.
XX
PD   16-OCT-1997.
XX
PF   04-APR-1997;   97WO-US05683.
XX
PR   04-APR-1996;   96US-0627967.
XX
PA   (HYBR-) HYBRIDON INC.
XX
PI   Agrawal S,  Kandimalla ER;
XX
DR   WPI; 1997-512714/47.
XX
PT   Anti:sense oligo:nucleotide compositions have at least 2 cooperative
PT   oligo:nucleotide(s) having a targeting and a dimerisation region -
PT   useful for inhibition of target nucleic acid expression
XX
PS   Disclosure; Page 26; 91pp; English.
XX
CC   The present oligonucleotide is an antisense oligonucleotide that binds
CC   to part of the gag gene of Human immunodeficiency virus type 1 (HIV-1).
CC   The present oligonucleotide has an extended sequence at the 5' end of
CC   the binding sequence which forms a duplex stem with the corresponding
CC   antisense dimerisation domain of AAT94871, when the 2 oligonucleotides
CC   bind to adjacent sites on the target sequence. The stability of binding
CC   was found to increase with the number of bases in the dimerisation
CC   domain. The oligonucleotides are used to exemplify the method of the
CC   invention. This method comprises two synthetic cooperative
CC   oligonucleotides, where each oligonucleotide comprises a region
CC   complementary to one of tandem, non-overlapping regions of a target
CC   nucleic acid, and a dimerisation domain at a terminus of each of the
CC   oligonucleotides. The dimerisation domains of the oligonucleotides are
CC   complementary to each other. The target nucleic acid is an mRNA, a
CC   single-stranded viral DNA, or a single-stranded viral RNA. The synthetic
CC   oligonucleotides can interact cooperatively to provide improved
CC   affinity, specificity, and biological activity as antisense molecules.
CC   The compositions are used for inhibiting the expression of target
CC   nucleic acids. They can be used for the treatment of viral infections,
CC   e.g. influenza (AAV04801-17) or HIV (AAT94853-92) infection. They can
CC   also be used for the detection and study of target nucleic acids.
XX
SQ   Sequence 17 BP; 0 A; 8 C; 3 G; 6 T; 0 other;

Alignment Scores:
Pred. No.:              1.14e+03        Length:         17
```

```
Score:                  15.00          Matches:       3
Percent Similarity:     60.00%         Conservative:  0
Best Local Similarity:  60.00%         Mismatches:    2
Query Match:            75.00%         Indels:        0
DB:                     18             Gaps:          0

US-09-726-470A-2 (1-8) x AAT94867 (1-17)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      3 CGGTCTCTCTCCTTC 17
```

RESULT 7
AAT47720/c
ID AAT47720 standard; DNA; 18 BP.
XX
AC AAT47720;
XX
DT 20-MAY-1997 (first entry)
XX
DE Mouse bone morphogenetic protein-4 gene RT-PCR primer 1.
XX
KW Osteogenic agent; bone morphogenetic protein-4; BMP-4;
KW growth factor; osteoblast; promoter; osteoporosis; fracture repair;
KW osteoblastic metastasis; osteosclerosis; therapy; primer; PCR;
KW polymerase chain reaction; ss.
XX
OS Synthetic.
XX
PN WO9638590-A1.
XX
PD 05-DEC-1996.
XX
PF 31-MAY-1996; 96WO-US08197.
XX
PR 02-JUN-1995; 95US-0458434.
XX
PA (OSTE-) OSTEOSCREEN INC.
XX
PI Feng JQ, Ghosh-Choudhury N, Harris SE, Mundy GR;
XX
DR WPI; 1997-034396/03.
XX
PT System for identifying osteogenic agents that induce prodn. of bone
PT morphogenetic protein - is cell contg. reporter gene under control of
PT BMP gene promoter, also new promoters of BMP-2 and -4 and related
PT vectors and cells
XX
PS Example 1; Page 11; 76pp; English.
XX
CC Primer 1 (AAT47720) corresponds to the 3' region of exon 1A of the
CC mouse bone morphogenetic protein-4 (BMP-4) gene (see also AAT47712).
CC It was used with primer 3 (AAT47721), corresponding to a 5' region of
CC exon 3, to generate an exon 1A-2-3 spliced PCR product. Primer 2
CC (AAT47722) corresponding to a 3' region of exon 1B and primer 3 were
CC used to generate exon 1B-2-3 spliced PCR products. Foetal rat
CC calvarial (FRC) cell total RNA was used as template. The results
CC indicated that FRC osteoblasts produce transcripts with either
CC exon 1A or 1B, but not both. 1A transcripts were 10-15 times more
CC abundant in primary bone cells.
XX
SQ Sequence 18 BP; 6 A; 3 C; 9 G; 0 U; 0 other;

```
Alignment Scores:
Pred. No.:              1.21e+03       Length:        18
Score:                  15.00          Matches:       3
Percent Similarity:     60.00%         Conservative:  0
Best Local Similarity:  60.00%         Mismatches:    2
Query Match:            75.00%         Indels:        0
DB:                     18             Gaps:          0

US-09-726-470A-2 (1-8) x AAT47720 (1-18)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     15 CGCGCTCTTGCCTTC 1
```

RESULT 8
AAZ44143/c
ID AAZ44143 standard; DNA; 18 BP.
XX
AC AAZ44143;
XX
DT 24-MAR-2000 (first entry)
XX
DE Human EGR-1 DNA antisense primer #24165.
XX
KW EGR-1; early growth response 1; antisense; inhibition; human; primer;
KW anti-inflammatory; cytostatic; antiviral; detection; diagnosis;
KW viral infection; inflammation; tumor; ss.
XX
OS Homo sapiens.
XX
PN US6008048-A.
XX
PD 28-DEC-1999.
XX
PF 04-DEC-1998; 98US-0205921.
XX
PR 04-DEC-1998; 98US-0205921.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Monia BP, Cowsert LM;
XX
DR WPI; 2000-096375/08.
XX
PT Antisense oligonucleotides that inhibit expression of human early
PT growth response-1, useful for diagnosis, treatment and prevention of
PT tumors, inflammation and infection -
XX
PS Example 15; Column 37-38; 31pp; English.
XX
CC This invention describes novel antisense oligonucleotides (I) capable of
CC inhibiting expression of human EGR-1 (early growth response-1). The
CC products of the invention have anti-inflammatory, cytostatic and
CC antiviral activity. (I) was tested for its effects on EGR-1 mRNA levels
CC by real-time polymerase chain reaction (PCR), results indicated that 60%
CC inhibition was achieved. When (I) was modified by 2'-O-methoxyethyl
CC substitution of the first 4 and last 4 residues, and by replacing any C
CC in these flanking regions with 5-methyl-C, the degree of inhibition was
CC increased to 71%. (I) is used to inhibit expression of EGR-1 in cells
CC and tissues in vitro, for research or diagnosis, e.g. detecting EGR-1
CC encoding nucleic acid. (I) may also be used to treat or prevent
CC EGR-1-associated diseases, particularly viral infections, inflammation
CC and tumors. AAZ44124-Z44169 represent antisense primers used to inhibit
CC the human EGR-1 protein.
XX
SQ Sequence 18 BP; 5 A; 2 C; 6 G; 5 T; 0 other;

```
Alignment Scores:
Pred. No.:              1.21e+03       Length:        18
Score:                  15.00          Matches:       3
Percent Similarity:     60.00%         Conservative:  0
Best Local Similarity:  60.00%         Mismatches:    2
Query Match:            75.00%         Indels:        0
DB:                     21             Gaps:          0

US-09-726-470A-2 (1-8) x AAZ44143 (1-18)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     17 CGAACACTGACATTT 3
```

RESULT 9
AAT41998

```
ID   AAT41998 standard; cDNA; 19 BP.
XX
AC   AAT41998;
XX
DT   24-JUN-1997  (first entry)
XX
DE   HIV-1 gag binding oligonucleotide used as antisense HIV inhibitor.
XX
KW   Co-operative binding; duplex; antisense inhibition; target sequence;
KW   human immunodeficiency virus; HIV; dimerisation domain; T structure;
KW   gene function; ss.
XX
OS   Synthetic.
XX
PN   WO9632474-A1.
XX
PD   17-OCT-1996.
XX
PF   04-APR-1996;   96WO-US04605.
XX
PR   12-APR-1995;   95US-0420672.
XX
PA   (HYBR-) HYBRIDON INC.
XX
PI   Agrawal S,  Kandimalla ER;
XX
DR   WPI; 1996-477125/47.
XX
PT   Compsn. contg. at least two co-operative oligo:nucleotide(s)
PT   complementary to a target sequence  - and with mutually
PT   complementary dimerisation domains, for use as antisense inhibitors
PT   of HIV and influenza virus
XX
PS   Disclosure; Page 46; 84pp; English.
XX
CC   AAT41998 is an oligonucleotide (ON) that binds to HIV-1 mRNA and/or
CC   DNA and act as antisense inhibitors of HIV-1 gene expression. Such
CC   ONs are preferably used as duplexes, i.e. a first ON has a region
CC   that binds to an HIV-1 target sequence in a 5'-3' direction and a
CC   second region complementary to a second ON which has a first region
CC   which binds to the same target HIV-1 sequence but in a 3'-5' direction
CC   and a second region complementary to the first ON. Both ONs bind to the
CC   HIV target sequence up until a certain point along the target sequence,
CC   where the two binding ONs are in close proximity and the remainder of
CC   the binding ONs bind to each other. The duplex/target site complex forms
CC   a T-shaped structure, inhibiting expression of HIV-1 nucleic acid. Also
CC   co-operative ONs, when labelled, can be used to identify specific
CC   bacteria or viruses in cell cultures; to study function of specific
CC   genes as an alternative to use of 'knock out' animals. Co-operative ONs
CC   have improved affinity and sequence specificity, reduced toxicity and
CC   better antisense activity compared with single, longer ONs. Insertion
CC   of dimerisation domains into antisense ONs (i.e. sequences hybridising
CC   to a second antisense ON) provides a more stable complex.
XX
SQ   Sequence 19 BP; 0 A; 9 C; 4 G; 6 T; 0 other;

Alignment Scores:
Pred. No.:              1.28e+03        Length:         19
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     17              Gaps:           0

US-09-726-470A-2 (1-8) x AAT41998 (1-19)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      5 CGGTCTCTCTCCTTC 19

RESULT 10
AAT94890
ID   AAT94890 standard; cDNA; 19 BP.
XX
AC   AAT94890;
XX
DT   22-APR-1998  (first entry)
XX
DE   HIV-1 gag gene antisense oligonucleotide.
XX
KW   gag gene; initiation codon region; target region; dimerisation domain;
KW   synthetic cooperative oligonucleotide; affinity; specificity;
KW   antisense molecule; treatment; viral infection; influenza; HIV; ss.
XX
OS   Synthetic.
OS   Human immunodeficiency virus type 1.
XX
FH   Key             Location/Qualifiers
FT   misc_binding    1..7
FT                   /*tag=  a
FT                   /note= "dimerisation domain which hybridises to
FT                          nucleotides 10-16 of AAT94872"
FT   misc_binding    8..19
FT                   /*tag=  b
FT                   /note= "binds to HIV-1 gag target AAT94853"
XX
PN   WO9738097-A1.
XX
PD   16-OCT-1997.
XX
PF   04-APR-1997;   97WO-US05683.
XX
PR   04-APR-1996;   96US-0627967.
XX
PA   (HYBR-) HYBRIDON INC.
XX
PI   Agrawal S,  Kandimalla ER;
XX
DR   WPI; 1997-512714/47.
XX
PT   Anti:sense oligo:nucleotide compositions have at least 2 cooperative
PT   oligo:nucleotide(s) having a targeting and a dimerisation region -
PT   useful for inhibition of target nucleic acid expression
XX
PS   Disclosure; Page 26; 91pp; English.
XX
CC   The present oligonucleotide is an antisense oligonucleotide that binds
CC   to part of the gag gene of Human immunodeficiency virus type 1 (HIV-1).
CC   The present oligonucleotide has an extended sequence at the 5' end of
CC   the binding sequence which forms a duplex stem with the corresponding
CC   antisense dimerisation domain of AAT94872, when the 2 oligonucleotides
CC   bind to adjacent sites on the target sequence. The stability of binding
CC   was found to increase with the number of bases in the dimerisation
CC   domain. The oligonucleotides are used to exemplify the method of the
CC   invention. This method comprises two synthetic cooperative
CC   oligonucleotides, where each oligonucleotide comprises a region
CC   complementary to one of tandem, non-overlapping regions of a target
CC   nucleic acid, and a dimerisation domain at a terminus of each of the
CC   oligonucleotides. The dimerisation domains of the oligonucleotides are
CC   complementary to each other. The target nucleic acid is an mRNA, a
CC   single-stranded viral DNA, or a single-stranded viral RNA. The synthetic
CC   oligonucleotides can interact cooperatively to provide improved
CC   affinity, specificity, and biological activity as antisense molecules.
CC   The compositions are used for inhibiting the expression of target
CC   nucleic acids. They can be used for the treatment of viral infections,
CC   e.g. influenza (AAV04801-17) or HIV (AAT94853-92) infection. They can
CC   also be used for the detection and study of target nucleic acids.
XX
SQ   Sequence 19 BP; 0 A; 9 C; 4 G; 6 T; 0 other;

Alignment Scores:
Pred. No.:              1.28e+03        Length:         19
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
```

DB: 18 Gaps: 0

US-09-726-470A-2 (1-8) x AAT94890 (1-19)

```
Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      5 CGGTCTCTCTCCTTC 19
```

RESULT 11
AAV40304
ID AAV40304 standard; DNA; 20 BP.
XX
AC AAV40304;
XX
DT 14-OCT-1998 (first entry)
XX
DE Maize oligonucleotide marker S01F.
XX
KW Maize; marker; probe; PCR primer; polymorphism; vegetal sequence;
KW polymorphic site; corn; graminae species; ss.
XX
OS Synthetic.
OS Zea sp.
XX
PN WO9830717-A2.
XX
PD 16-JUL-1998.
XX
PF 02-DEC-1997; 97WO-EP07134.
XX
PR 02-DEC-1996; 96US-0032069.
XX
PA (BIOC-) BIOCEM SA.
XX
PI Murigneux A;
XX
DR WPI; 1998-399160/34.
XX
PT Vegetal sequences including single nucleotide polymorphism - useful,
PT e.g. to determine polymorphisms in plants, determine strain in plant
PT breeding and to correlate polymorphisms with phenotypic traits
XX
PS Example 2; Page 9; 32pp; English.
XX
CC The present invention describes a nucleic acid segment comprising at
CC least 10 contiguous nucleotides from a vegetal sequence including a
CC polymorphic site which is a single nucleotide polymorphism (SNP), or the
CC complement of the segment. Also described are: (1) an allele-specific
CC oligonucleotides hybridising to segment, or their complements, and (2) a
CC method of analysing nucleic acids from a subject, by determining if a
CC base is occupying any one (or a set) of polymorphic sites in 261
CC sequences derived from six maize lines (see AAV47701 to AAV47961). The
CC segments are useful in fingerprint analysis in plants to determine which
CC polymorphisms are present, which strain a plant belongs to and to
CC distinguish between strains. The polymorphisms may correlate with
CC phenotypic traits (e.g. plant growth rate or crop yield), and the
CC segments are useful to determine the presence/absence of specific
CC polymorphisms correlating with the existence/absence of particular
CC traits. The segments are also useful in marker assisted back-cross
CC techniques to select plants with a higher percentage of recurrent parent
CC in a back-cross population. Segments incorporate SNPs which occur more
CC frequently than other polymorphism types and are therefore more likely
CC to be located close to genetic loci of interest; different forms of
CC characterised SNPs are also often easier to detect than other
CC polymorphism types. AAV40304 to AAV40369 are used in an example from the
CC present invention as markers and PCR primers.
XX
SQ Sequence 20 BP; 4 A; 7 C; 4 G; 5 T; 0 other;

Alignment Scores:

Pred. No.:	1.35e+03	Length:	20
Score:	15.00	Matches:	3
Percent Similarity:	60.00%	Conservative:	0
Best Local Similarity:	60.00%	Mismatches:	2
Query Match:	75.00%	Indels:	0
DB:	19	Gaps:	0

US-09-726-470A-2 (1-8) x AAV40304 (1-20)

```
Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      5 CGCACATTAGCTTTC 19
```

RESULT 12
AAZ05921
ID AAZ05921 standard; DNA; 20 BP.
XX
AC AAZ05921;
XX
DT 07-OCT-1999 (first entry)
XX
DE PCR primer used to amplify an ORF of Chlamydia trachomatis.
XX
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW nongonococcal uretritis; epidymitis; cervicitis; salpingitis; PCR primer;
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX
OS Synthetic.
OS Chlamydia trachomatis.
XX
PN WO9928475-A2.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98WO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GEST) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
XX
PS Disclosure; Page 1810; 1755pp; English.
XX
CC PCR primers AAZ01426-Z06209 were used to amplify open reading frames
CC (ORFs) of the genome of Chlamydia trachomatis (see AAZ01425). These ORFs
CC encode polypeptides (see AAY36754-Y37949) which can be used as vaccines
CC against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye
CC diseases such as conventional trachoma, nonendemic trachoma,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal uretritis, epidymitis, cervicitis, salpingitis,
CC perihepatitis, bartholinitis; pneumopathy in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the
CC invention may be of use in treating these diseases.
XX
SQ Sequence 20 BP; 2 A; 8 C; 2 G; 8 T; 0 other;

Alignment Scores:

Pred. No.:	1.35e+03	Length:	20
Score:	15.00	Matches:	3
Percent Similarity:	60.00%	Conservative:	0
Best Local Similarity:	60.00%	Mismatches:	2
Query Match:	75.00%	Indels:	0
DB:	20	Gaps:	0

US-09-726-470A-2 (1-8) x AAZ05921 (1-20)

```
Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      1 CGATCTCTCTCCTTT 15

RESULT 13
AAZ05924
ID   AAZ05924 standard; DNA; 20 BP.
XX
AC   AAZ05924;
XX
DT   07-OCT-1999  (first entry)
XX
DE   PCR primer used to amplify an ORF of Chlamydia trachomatis.
XX
KW   Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW   paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW   nongonococcal uretritis; epidymitis; cervicitis; salpingitis; PCR primer;
KW   bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX
OS   Synthetic.
OS   Chlamydia trachomatis.
XX
PN   WO9928475-A2.
XX
PD   10-JUN-1999.
XX
PF   27-NOV-1998;   98WO-IB01939.
XX
PR   04-NOV-1998;   98US-0107077.
PR   28-NOV-1997;   97FR-0015041.
PR   17-DEC-1997;   97FR-0016034.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-371125/31.
XX
PT   Genome sequence of Chlamydia trachomatis
XX
PS   Disclosure; Page 1810; 1755pp; English.
XX
CC   PCR primers AAZ01426-Z06209 were used to amplify open reading frames
CC   (ORFs) of the genome of Chlamydia trachomatis (see AAZ01425). These ORFs
CC   encode polypeptides (see AAY36754-Y37949) which can be used as vaccines
CC   against Chlamydia trachomatis. Antisense and ribozyme sequences
CC   can also be used to control growth of the microorganism. Chlamydia
CC   trachomatis is responsible for a large number of diseases, e.g. eye
CC   diseases such as conventional trachoma, nonendemic trachoma,
CC   paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC   nongonococcal uretritis, epidymitis, cervicitis, salpingitis,
CC   perihepatitis, bartholinitis; pneumopathy in breast feeding infants;
CC   and venereal lymphogranulomatosis. The polypeptides of the
CC   invention may be of use in treating these diseases.
XX
SQ   Sequence 20 BP; 2 A; 8 C; 2 G; 8 T; 0 other;

Alignment Scores:
Pred. No.:              1.35e+03        Length:         20
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     20              Gaps:           0

US-09-726-470A-2 (1-8) x AAZ05924 (1-20)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      1 CGATCTCTCTCCTTT 15

RESULT 14
AAX97011
ID   AAX97011 standard; DNA; 20 BP.
XX
AC   AAX97011;
XX
DT   13-SEP-1999  (first entry)
XX
DE   PCR primer used to amplify an ORF of Chlamydia pneumoniae.
XX
KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW   sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW   vaccine; neutralising epitope; PCR primer; ss.
XX
OS   Synthetic.
OS   Chlamydia pneumoniae.
XX
PN   WO9927105-A2.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;   98WO-IB01890.
XX
PR   04-NOV-1998;   98US-0107078.
PR   21-NOV-1997;   97FR-0014673.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-357842/30.
XX
PT   Genome sequence of Chlamydia pneumoniae
XX
PS   Page 1871; Disclosure; 1912pp; English.
XX
CC   AAX91991-X97517 represent PCR primers used to amplify open reading
CC   frames and other nucleic acid sequences from the genome of
CC   Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
CC   disease such as pneumonia and bronchitis and is thought to be a
CC   contributing factor in heart disease, sarcoidosis, sinusitis, purulent
CC   otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
CC   by the open reading frames of the C. pneumoniae genome (see AAY34584-
CC   AAY35879) can be used in immunogenic compositions as vaccines. Vectors
CC   containing C. pneumoniae nucleotides sequences can also be used as
CC   immunogenic compositions, especially where the vector directs the
CC   expression of a neutralising epitope of C. pneumoniae.
XX
SQ   Sequence 20 BP; 3 A; 8 C; 4 G; 5 T; 0 other;

Alignment Scores:
Pred. No.:              1.35e+03        Length:         20
Score:                  15.00           Matches:        3
Percent Similarity:     60.00%          Conservative:   0
Best Local Similarity:  60.00%          Mismatches:     2
Query Match:            75.00%          Indels:         0
DB:                     20              Gaps:           0

US-09-726-470A-2 (1-8) x AAX97011 (1-20)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      1 CGGGCACTTTCCTTC 15

RESULT 15
AAX94959/c
ID   AAX94959 standard; DNA; 20 BP.
XX
AC   AAX94959;
XX
DT   13-SEP-1999  (first entry)
XX
DE   PCR primer used to amplify an ORF of Chlamydia pneumoniae.
XX
KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
```

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope; PCR primer; ss.
XX
OS Synthetic.
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Page 1710; Disclosure; 1912pp; English.
XX
CC AAX91991-X97517 represent PCR primers used to amplify open reading
CC frames and other nucleic acid sequences from the genome of
CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
CC disease such as pneumonia and bronchitis and is thought to be a
CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors
CC containing C. pneumoniae nucleotides sequences can also be used as
CC immunogenic compositions, especially where the vector directs the
CC expression of a neutralising epitope of C. pneumoniae.
XX
SQ Sequence 20 BP; 4 A; 2 C; 8 G; 6 T; 0 other;

Alignment Scores:

Pred. No.:	1.35e+03	Length:	20
Score:	15.00	Matches:	3
Percent Similarity:	60.00%	Conservative:	0
Best Local Similarity:	60.00%	Mismatches:	2
Query Match:	75.00%	Indels:	0
DB:	20	Gaps:	0

US-09-726-470A-2 (1-8) x AAX94959 (1-20)

```
Qy       4 Arg***Leu***Phe 8
           |||   |||   |||
Db      15 AGAACCCTCGCCTTC 1
```

Search completed: December 14, 2002, 16:00:05
Job time : 221 secs

GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2002, 15:50:15 ; Search time 1607 Seconds
(without alignments)
144.880 Million cell updates/sec

Title: US-09-726-470A-2
Perfect score: 20
Sequence: 1 XXXRXLXF 8

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09726470/runat_10122002_090717_4972/app_query.fasta_1.398
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09726470_@CGN_1_1_3637_@runat_10122002_090717_4972 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		15	75.0	15	6	AR206330	AR206330 Sequence
2		15	75.0	16	6	AR206331	AR206331 Sequence
3		15	75.0	17	6	AR206332	AR206332 Sequence
4	c	15	75.0	18	6	AR096637	AR096637 Sequence
5	c	15	75.0	18	6	AR100893	AR100893 Sequence
6		15	75.0	19	6	AR206333	AR206333 Sequence
7		15	75.0	20	6	AR095074	AR095074 Sequence
8		15	75.0	20	6	AR100389	AR100389 Sequence
9		15	75.0	20	6	AR117594	AR117594 Sequence
10		15	75.0	20	6	AR150044	AR150044 Sequence
11	c	15	75.0	20	6	AX036966	AX036966 Sequence
12		15	75.0	20	6	AX036972	AX036972 Sequence
13	c	15	75.0	21	6	AR202588	AR202588 Sequence
14	c	15	75.0	21	6	AX298463	AX298463 Sequence
15		15	75.0	21	6	E10426	E10426 Primer. 9/1
16	c	15	75.0	22	6	AR072181	AR072181 Sequence
17	c	15	75.0	22	6	AR086550	AR086550 Sequence
18		15	75.0	22	6	AR086551	AR086551 Sequence
19		15	75.0	23	6	AR207290	AR207290 Sequence
20	c	15	75.0	24	6	AX055922	AX055922 Sequence
21		15	75.0	24	6	AX252517	AX252517 Sequence
22		15	75.0	24	6	AX253133	AX253133 Sequence
23		15	75.0	24	6	AX446204	AX446204 Sequence
24		15	75.0	24	6	AX447514	AX447514 Sequence
25		15	75.0	24	6	I26953	I26953 Sequence 2
26		15	75.0	24	6	I28389	I28389 Sequence 2
27		15	75.0	24	6	I96081	I96081 Sequence 2
28		15	75.0	25	6	AX278996	AX278996 Sequence
29	c	15	75.0	26	6	AX351078	AX351078 Sequence
30	c	15	75.0	27	6	AX299932	AX299932 Sequence
31		15	75.0	27	6	AX466784	AX466784 Sequence
32		15	75.0	27	6	AX466814	AX466814 Sequence
33		15	75.0	27	6	E11109	E11109 Mutagenesis
34		15	75.0	27	6	E12128	E12128 PCR primer
35		15	75.0	27	6	I74625	I74625 Sequence 5
36		15	75.0	27	6	I91958	I91958 Sequence 6
37	c	15	75.0	29	6	AX461477	AX461477 Sequence
38		15	75.0	30	6	A08039	A08039 Oligonucleo
39	c	15	75.0	30	6	A14208	A14208 oligonucleo
40		15	75.0	30	6	AR004723	AR004723 Sequence
41		15	75.0	30	6	AR008209	AR008209 Sequence
42		15	75.0	30	6	AR034014	AR034014 Sequence
43		15	75.0	30	6	AR124023	AR124023 Sequence
44		15	75.0	30	6	AR136992	AR136992 Sequence
45		15	75.0	30	6	AX474209	AX474209 Sequence

ALIGNMENTS

RESULT 1

```
AR206330
LOCUS       AR206330                 15 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 10 from patent US 6372427.
ACCESSION   AR206330
VERSION     AR206330.1  GI:21504900
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 15)
  AUTHORS   Kandimalla,E.R. and Agrawal,S.
  TITLE     Cooperative oligonucleotides
  JOURNAL   Patent: US 6372427-A 10 16-APR-2002;
FEATURES             Location/Qualifiers
     source          1. .15
                     /organism="unknown"
BASE COUNT        0 a      7 c      2 g      6 t
ORIGIN

Alignment Scores:
Pred. No.:              662           Length:         15
Score:                  15.00         Matches:        3
Percent Similarity:     60.00%        Conservative:   0
Best Local Similarity:  60.00%        Mismatches:     2
Query Match:            75.00%        Indels:         0
DB:                     6             Gaps:           0

US-09-726-470A-2 (1-8) x AR206330 (1-15)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      1 CGGTCTCTCTCCTTC 15

RESULT 2
AR206331
LOCUS       AR206331                 16 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 11 from patent US 6372427.
ACCESSION   AR206331
VERSION     AR206331.1  GI:21504901
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 16)
  AUTHORS   Kandimalla,E.R. and Agrawal,S.
  TITLE     Cooperative oligonucleotides
  JOURNAL   Patent: US 6372427-A 11 16-APR-2002;
FEATURES             Location/Qualifiers
     source          1. .16
                     /organism="unknown"
BASE COUNT        0 a      8 c      2 g      6 t
ORIGIN

Alignment Scores:
Pred. No.:              706           Length:         16
Score:                  15.00         Matches:        3
Percent Similarity:     60.00%        Conservative:   0
Best Local Similarity:  60.00%        Mismatches:     2
Query Match:            75.00%        Indels:         0
DB:                     6             Gaps:           0

US-09-726-470A-2 (1-8) x AR206331 (1-16)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      2 CGGTCTCTCTCCTTC 16

RESULT 3
AR206332
LOCUS       AR206332                 17 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 12 from patent US 6372427.
ACCESSION   AR206332
VERSION     AR206332.1  GI:21504902
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 17)
  AUTHORS   Kandimalla,E.R. and Agrawal,S.
  TITLE     Cooperative oligonucleotides
  JOURNAL   Patent: US 6372427-A 12 16-APR-2002;
FEATURES             Location/Qualifiers
     source          1. .17
                     /organism="unknown"
BASE COUNT        0 a      8 c      3 g      6 t
ORIGIN

Alignment Scores:
Pred. No.:              750           Length:         17
Score:                  15.00         Matches:        3
Percent Similarity:     60.00%        Conservative:   0
Best Local Similarity:  60.00%        Mismatches:     2
Query Match:            75.00%        Indels:         0
DB:                     6             Gaps:           0

US-09-726-470A-2 (1-8) x AR206332 (1-17)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      3 CGGTCTCTCTCCTTC 17

RESULT 4
AR096637/c
LOCUS       AR096637                 18 bp    DNA     linear   PAT 08-SEP-2000
DEFINITION  Sequence 21 from patent US 6008048.
ACCESSION   AR096637
VERSION     AR096637.1  GI:10025610
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Monia,B.P. and Cowsert,L.M.
  TITLE     Antisense inhibition of EGR-1 expression
  JOURNAL   Patent: US 6008048-A 21 28-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unknown"
BASE COUNT        5 a      2 c      6 g      5 t
ORIGIN

Alignment Scores:
Pred. No.:              795           Length:         18
Score:                  15.00         Matches:        3
Percent Similarity:     60.00%        Conservative:   0
Best Local Similarity:  60.00%        Mismatches:     2
Query Match:            75.00%        Indels:         0
DB:                     6             Gaps:           0

US-09-726-470A-2 (1-8) x AR096637 (1-18)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     17 CGAACACTGACATTT 3

RESULT 5
AR100893/c
LOCUS       AR100893                 18 bp    DNA     linear   PAT 14-FEB-2001
DEFINITION  Sequence 11 from patent US 6083690.
ACCESSION   AR100893
VERSION     AR100893.1  GI:12811691
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 18)
```

```
  AUTHORS   Harris,S.E., Mundy,G.R., Ghosh-Choudhury,N. and Feng,J.Q.
  TITLE     Methods and compositions for identifying osteogenic agents
  JOURNAL   Patent: US 6083690-A 11 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unknown"
BASE COUNT        6 a      3 c      9 g      0 t
ORIGIN

Alignment Scores:
Pred. No.:              795           Length:        18
Score:                  15.00         Matches:       3
Percent Similarity:     60.00%        Conservative:  0
Best Local Similarity:  60.00%        Mismatches:    2
Query Match:            75.00%        Indels:        0
DB:                     6             Gaps:          0

US-09-726-470A-2 (1-8) x AR100893 (1-18)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     15 CGCGCTCTTGCCTTC 1

RESULT 6
AR206333
LOCUS       AR206333                19 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 13 from patent US 6372427.
ACCESSION   AR206333
VERSION     AR206333.1  GI:21504904
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 19)
  AUTHORS   Kandimalla,E.R. and Agrawal,S.
  TITLE     Cooperative oligonucleotides
  JOURNAL   Patent: US 6372427-A 13 16-APR-2002;
FEATURES             Location/Qualifiers
     source          1. .19
                     /organism="unknown"
BASE COUNT        0 a      9 c      4 g      6 t
ORIGIN

Alignment Scores:
Pred. No.:              839           Length:        19
Score:                  15.00         Matches:       3
Percent Similarity:     60.00%        Conservative:  0
Best Local Similarity:  60.00%        Mismatches:    2
Query Match:            75.00%        Indels:        0
DB:                     6             Gaps:          0

US-09-726-470A-2 (1-8) x AR206333 (1-19)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      5 CGGTCTCTCTCCTTC 19

RESULT 7
AR095074
LOCUS       AR095074                20 bp    DNA     linear   PAT 08-SEP-2000
DEFINITION  Sequence 27 from patent US 6001992.
ACCESSION   AR095074
VERSION     AR095074.1  GI:10022599
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Ackermann,E.J., Bennett,C.Frank., Dean,N.M. and Marcusson,E.G.
  TITLE     Antisense modulation of novel anti-apoptotic bcl-2-related proteins
  JOURNAL   Patent: US 6001992-A 27 14-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unknown"
BASE COUNT        4 a      6 c      4 g      6 t
ORIGIN

Alignment Scores:
Pred. No.:              883           Length:        20
Score:                  15.00         Matches:       3
Percent Similarity:     60.00%        Conservative:  0
Best Local Similarity:  60.00%        Mismatches:    2
Query Match:            75.00%        Indels:        0
DB:                     6             Gaps:          0

US-09-726-470A-2 (1-8) x AR095074 (1-20)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      4 AGGTCACTGGCATTC 18

RESULT 8
AR100389
LOCUS       AR100389                20 bp    DNA     linear   PAT 14-FEB-2001
DEFINITION  Sequence 120 from patent US 6080580.
ACCESSION   AR100389
VERSION     AR100389.1  GI:12810837
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Baker,B.F., Bennett,C.Frank., Butler,M.M. and Shanahan,W.R. Jr.
  TITLE     Antisense oligonucleotide modulation of tumor necrosis
            factor-.alpha. (TNF-.alpha.) expression
  JOURNAL   Patent: US 6080580-A 120 27-JUN-2000;
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unknown"
BASE COUNT        3 a      5 c      4 g      8 t
ORIGIN

Alignment Scores:
Pred. No.:              883           Length:        20
Score:                  15.00         Matches:       3
Percent Similarity:     60.00%        Conservative:  0
Best Local Similarity:  60.00%        Mismatches:    2
Query Match:            75.00%        Indels:        0
DB:                     6             Gaps:          0

US-09-726-470A-2 (1-8) x AR100389 (1-20)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      1 AGAGCTCTGTCTTTT 15

RESULT 9
AR117594
LOCUS       AR117594                20 bp    DNA     linear   PAT 16-MAY-2001
DEFINITION  Sequence 86 from patent US 6140124.
ACCESSION   AR117594
VERSION     AR117594.1  GI:14098500
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Monia,B.P., Gaarde,W.A., Nero,P.S. and McKay,R.
  TITLE     Antisense modulation of P38 mitogen activated protein kinase
            expression
  JOURNAL   Patent: US 6140124-A 86 31-OCT-2000;
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unknown"
BASE COUNT        4 a      6 c      3 g      7 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                883           Length:        20
Score:                    15.00         Matches:       3
Percent Similarity:       60.00%        Conservative:  0
Best Local Similarity:    60.00%        Mismatches:    2
Query Match:              75.00%        Indels:        0
DB:                       6             Gaps:          0

US-09-726-470A-2 (1-8) x AR117594 (1-20)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      3 CGTAGCCTGTCATTT 17

RESULT 10
AR150044
LOCUS       AR150044                  20 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 120 from patent US 6228642.
ACCESSION   AR150044
VERSION     AR150044.1  GI:15114635
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Baker,B.F., Bennett,C.Frank., Butler,M.M. and Shanahan,W.R. Jr.
  TITLE     Antisense oligonucleotide modulation of tumor necrosis
            factor-(.alpha.) (TNF-.alpha.) expression
  JOURNAL   Patent: US 6228642-A 120 08-MAY-2001;
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unknown"
BASE COUNT        3 a      5 c      4 g      8 t
ORIGIN

Alignment Scores:
Pred. No.:                883           Length:        20
Score:                    15.00         Matches:       3
Percent Similarity:       60.00%        Conservative:  0
Best Local Similarity:    60.00%        Mismatches:    2
Query Match:              75.00%        Indels:        0
DB:                       6             Gaps:          0

US-09-726-470A-2 (1-8) x AR150044 (1-20)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      1 AGAGCTCTGTCTTTT 15

RESULT 11
AX036966/c
LOCUS       AX036966                  20 bp    DNA     linear   PAT 16-NOV-2000
DEFINITION  Sequence 23 from Patent FR2790955.
ACCESSION   AX036966
VERSION     AX036966.1  GI:11226394
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Carpentier,A.
  JOURNAL   Patent: FR 2790955-A 23 22-SEP-2000;
            ASSIST PUBL HOPITAUX DE PARIS (FR)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="oligodesoxynucleotide"
BASE COUNT        9 a      2 c      4 g      5 t
ORIGIN

Alignment Scores:
Pred. No.:                883           Length:        20
Score:                    15.00         Matches:       3
Percent Similarity:       60.00%        Conservative:  0
Best Local Similarity:    60.00%        Mismatches:    2
Query Match:              75.00%        Indels:        0
DB:                       6             Gaps:          0

US-09-726-470A-2 (1-8) x AX036966 (1-20)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     16 CGTTCATTAACGTTC 2

RESULT 12
AX036972
LOCUS       AX036972                  20 bp    DNA     linear   PAT 16-NOV-2000
DEFINITION  Sequence 29 from Patent FR2790955.
ACCESSION   AX036972
VERSION     AX036972.1  GI:11226400
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Carpentier,A.
  JOURNAL   Patent: FR 2790955-A 29 22-SEP-2000;
            ASSIST PUBL HOPITAUX DE PARIS (FR)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="oligodesoxynucleotide"
BASE COUNT        4 a      4 c      2 g     10 t
ORIGIN

Alignment Scores:
Pred. No.:                883           Length:        20
Score:                    15.00         Matches:       3
Percent Similarity:       60.00%        Conservative:  0
Best Local Similarity:    60.00%        Mismatches:    2
Query Match:              75.00%        Indels:        0
DB:                       6             Gaps:          0

US-09-726-470A-2 (1-8) x AX036972 (1-20)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      5 CGTTCTTTAACGTTC 19

RESULT 13
AR202588/c
LOCUS       AR202588                  21 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 3 from patent US 6365126.
ACCESSION   AR202588
VERSION     AR202588.1  GI:21498757
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Zhong,Y., Guo,H.-F. and Tong,J.
  TITLE     Learning and short term memory defects with Neurofibromatosis 1
            (NF1) expression
  JOURNAL   Patent: US 6365126-A 3 02-APR-2002;
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unknown"
BASE COUNT        8 a      7 c      4 g      2 t
ORIGIN

Alignment Scores:
Pred. No.:                927           Length:        21
Score:                    15.00         Matches:       3
```

```
Percent Similarity:     60.00%          Conservative:  0
Best Local Similarity:  60.00%          Mismatches:    2
Query Match:            75.00%          Indels:        0
DB:                     6               Gaps:          0

US-09-726-470A-2 (1-8) x AR202588 (1-21)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     20 CGTGCTCTGAGCTTT 6

RESULT 14
AX298463/c
LOCUS       AX298463                  21 bp    DNA     linear   PAT 26-NOV-2001
DEFINITION  Sequence 97 from Patent WO0183749.
ACCESSION   AX298463
VERSION     AX298463.1  GI:17128453
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Bachmanov,A.A., Beauchamp,G.K., Chatterjee,A., de Jong,P.J., Li,S.,
            Li,X., Ohmen,J.D., Reed,D.R., Ross,D. and Tordoff,M.G.
  TITLE     Gene and sequence variation associated with sensing carbohydrate
            compounds and other sweeteners
  JOURNAL   Patent: WO 0183749-A 97 08-NOV-2001;
            WARNER-LAMBERT COMPANY (US) ; The Monell Chemical Senses Center
            (US)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
BASE COUNT        8 a      5 c      7 g      1 t
ORIGIN

Alignment Scores:
Pred. No.:              927             Length:        21
Score:                  15.00           Matches:       3
Percent Similarity:     60.00%          Conservative:  0
Best Local Similarity:  60.00%          Mismatches:    2
Query Match:            75.00%          Indels:        0
DB:                     6               Gaps:          0

US-09-726-470A-2 (1-8) x AX298463 (1-21)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db     16 CGTAGCCTGGCTTTC 2

RESULT 15
E10426
LOCUS       E10426                    21 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  Primer.
ACCESSION   E10426
VERSION     E10426.1  GI:22027259
KEYWORDS    JP 1995327681-A/3.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Yoshigi,H. and Maehane,H.
  TITLE     RECOMBINANT BETA-AMYLASE IMPROVED IN THERMAL STABILITY
  JOURNAL   Patent: JP 1995327681-A 3 19-DEC-1995;
            SAPPORO BREWERIES LTD
COMMENT     OS   None
            OC   Artificial sequences.
            PN   JP 1995327681-A/3
            PD   19-DEC-1995
            PF   08-JUN-1994 JP 1994126151
            PI   YOSHIGI HISAHIRO, MAEHANE HIDEO
            PC   C12N15/09,C12N1/21,C12N9/26,(C12N1/21,C12R1:19),(C12N9/26, PC
             C12R1:19);
            CC   strandedness: Single;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .21
            FT                   /organism='Artificial sequences'.
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        7 a      5 c      4 g      5 t
ORIGIN

Alignment Scores:
Pred. No.:              927             Length:        21
Score:                  15.00           Matches:       3
Percent Similarity:     60.00%          Conservative:  0
Best Local Similarity:  60.00%          Mismatches:    2
Query Match:            75.00%          Indels:        0
DB:                     6               Gaps:          0

US-09-726-470A-2 (1-8) x E10426 (1-21)

Qy      4 Arg***Leu***Phe 8
          |||   |||   |||
Db      6 AGATCGCTGGCATTC 20


Search completed: December 14, 2002, 16:53:54
Job time : 1610 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: December 14, 2002, 13:14:29 ; Search time 57 Seconds
(without alignments)
18.702 Million cell updates/sec

Title: US-09-726-470A-2
Perfect score: 20
Sequence: 1 XXXRXLXF 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	15	75.0	15	14	AAR30969	Jacalin fragment w
2	15	75.0	21	15	AAR51378	J alpha sequence (
3	15	75.0	27	21	AAY78753	Hypervariable regi
4	15	75.0	27	23	AAE18094	Biotinylated hepat
5	15	75.0	28	21	AAB38998	Human secreted pep
6	15	75.0	30	20	AAY02875	Fragment of human
7	15	75.0	31	18	AAW45039	Immunomodulatory p
8	15	75.0	31	20	AAY09486	Immunoactive pepti
9	15	75.0	33	21	AAB44770	Human secreted pro
10	15	75.0	33	22	AAM60971	Human brain expres
11	15	75.0	33	22	AAM32187	Peptide #6224 enco
12	15	75.0	33	23	ABG41720	Human peptide enco
13	15	75.0	33	23	ABG43538	Human peptide enco
14	15	75.0	35	21	AAB44347	Human secreted pro
15	15	75.0	36	21	AAB63066	Human secreted pro
16	15	75.0	38	23	ABG68830	Cytochrome P450 3A
17	15	75.0	40	21	AAG15201	Arabidopsis thalia
18	15	75.0	43	20	AAY25813	Human secreted pro
19	15	75.0	45	22	AAU86594	Novel human connec
20	15	75.0	49	22	ABG26048	Novel human diagno
21	15	75.0	50	22	AAM86216	Human immune/haema
22	15	75.0	53	23	ABG66860	Human prostate spe
23	15	75.0	56	22	AAU67825	Propionibacterium
24	15	75.0	56	23	ABG63221	Human prostate spe
25	15	75.0	56	23	ABP07362	Human ORFX protein
26	15	75.0	57	23	ABP34657	Human ORF3630 prot
27	15	75.0	58	21	AAY58437	Staphylococcus aur
28	15	75.0	58	22	ABG26045	Novel human diagno
29	15	75.0	58	22	AAM38400	Peptide #12437 enc
30	15	75.0	58	22	AAB69510	Staphylococcus aur
31	15	75.0	58	23	ABG47169	Human peptide enco
32	15	75.0	59	10	AAP91536	Modified region 48
33	15	75.0	60	22	AAG99889	ERA binding domain
34	15	75.0	61	20	AAY59684	Secreted protein 1
35	15	75.0	62	23	ABP33448	Human ORF2421 prot
36	15	75.0	63	22	AAU22675	Novel human colon
37	15	75.0	63	22	AAM92859	Human digestive sy
38	15	75.0	63	23	ABP06349	Human ORFX protein
39	15	75.0	63	23	ABB49639	Listeria monocytog
40	15	75.0	65	22	ABB17357	Human nervous syst
41	15	75.0	65	22	AAM17297	Peptide #3731 enco
42	15	75.0	65	22	AAM04979	Peptide #3661 enco
43	15	75.0	65	23	ABP05290	Human ORFX protein
44	15	75.0	66	22	ABB11901	Human cytokine-lik
45	15	75.0	67	21	AAY52239	E. coli ygjD prote

ALIGNMENTS

RESULT 1
AAR30969
ID AAR30969 standard; peptide; 15 AA.
XX
AC AAR30969;
XX
DT 07-MAY-1993 (first entry)
XX
DE Jacalin fragment which interacts with CD4 receptor.
XX
KW Jackfruit; human immunodeficiency virus; HIV-1; gp120;
KW mitogenic lectin.
XX
OS Artocarpus heterophyllus.
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "uncharged hydrophilic amino acid
FT - opt. absent"
XX
PN WO9222574-A.
XX
PD 23-DEC-1992.
XX
PF 05-JUN-1992; 92WO-FR00510.
XX
PR 10-JUN-1991; 91FR-0007041.
PR 31-JAN-1992; 92FR-0001127.
XX
PA (INRM) INSERM INST NAT SANTE & RECH MED.
XX
PI Corbeau P, Devaux C, Dornand J, Favero J, Nicolas M;
PI Liautard J;

XX
DR WPI; 1993-018076/02.
XX
PT Jacalin and its new peptide fragments for treating HIV -
PT interacting with the CD4 receptor and specifically preventing
PT infection of lymphocytes
XX
PS Claim 7; Page 14; 27pp; French.
XX
CC This fragment of Jacalin interacts with the CD4 receptor and is
CC homologous with the sequence of the HIV protein gp120. The peptide
CC and other peptides with biological activity equivalent to that of
CC Jacalin are useful in treatment of diseases caused by HIV. They
CC specifically inhibit infection of lymphocytes by HIV, do not affect
CC normal lymphocyte function and (unlike Jacalin itself) do not
CC agglutinate cells.
XX
SQ Sequence 15 AA;

Query Match 75.0%; Score 15; DB 14; Length 15;
Best Local Similarity 60.0%; Pred. No. 1.7e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 5 RSLTF 9

RESULT 2
AAR51378
ID AAR51378 standard; Protein; 21 AA.
XX
AC AAR51378;
XX
DT 20-OCT-1994 (first entry)
XX
DE J alpha sequence (Val2.1/JaA?? usage).
XX
KW Rheumatoid arthritis; antibody; TCR; T cell receptor; lymphocyte;
KW expansion; complementary determining region; CDR3; antigen; MHC.
XX
OS Homo sapiens.
XX
PN WO9406823-A.
XX
PD 31-MAR-1994.
XX
PF 14-SEP-1993; 93WO-US08644.
XX
PR 14-SEP-1992; 92US-0943418.
XX
PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
XX
PI Brenner MB, Dersimonian H;
XX
DR WPI; 1994-118395/14.
XX
PT Treatment and prevention of rheumatoid arthritis - using peptides
PT derived from DQw2 or antibodies to DQw2 which block activation of
PT T lymphocytes
XX
PS Disclosure; Fig 3A; 57pp; English.
XX
CC To gain insight into the basis fo the Valpha12.1+T cell expansion
CC in rheumatoid arthritis, Valpha12.1 transcripts from positively
CC selected CD8+T cells were cloned and sequenced. In each of the
CC three patients analysed, distinct, repeated Valpha12.1 contg.
CC sequences corresp. to functional TCR alpha-chain transcripts were
CC identified. All of the repeated Valpha12.1+ T cell rearrangements
CC in the 3 patients analysed use either JalphaA1, JalphaA12 or
CC JalphaA6, each of which encodes a unique sequence at the 3' end
CC of the Jalpha gene segment. This short stretch of shared residues
CC (pro-tyr) is predicted to contribute (or is immediately adjacent)
CC to the third complementary determining region (CDR3) and thus may
CC play a role in antigen or MHC recognition.
XX
SQ Sequence 21 AA;

Query Match 75.0%; Score 15; DB 15; Length 21;
Best Local Similarity 60.0%; Pred. No. 2.4e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 6 RALTF 10

RESULT 3
AAY78753
ID AAY78753 standard; Peptide; 27 AA.
XX
AC AAY78753;
XX
DT 08-MAY-2000 (first entry)
XX
DE Hypervariable region 1 representative peptide sequence 9.
XX
KW Hepatitis C virus; envelope protein E2; hypervariable region 1; mimitope;
KW peptide library; treatment; prevent infection; antibody production.
XX
OS Hepatitis C virus.
XX
PN WO9960132-A1.
XX
PD 25-NOV-1999.
XX
PF 14-MAY-1999; 99WO-EP03344.
XX
PR 19-MAY-1998; 98GB-0010756.
XX
PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI Nicosia A, Lahm A, Tramontano A, Cortese R;
XX
DR WPI; 2000-126382/11.
XX
PT A new peptide library from hepatitis C virus, useful for production of
PT treatment for hepatitis C -
XX
PS Examples; Page 73; 126pp; English.
XX
CC This sequence represents a peptide from the library of the invention.
CC The invention relates to a library of peptides which have an
CC immunologically reactive epitope of the hypervariable region 1 (HVR1) of
CC envelope protein 2 (E2) of hepatitis C virus. The peptides contained in
CC the library correspond to formulae given in the specification (see
CC AAY78596-Y78598). This sequence is included in a selection of a
CC representative set of natural HVR1 sequences. The peptides can be used in
CC a method to select antibodies which react with the HVR1 of E2 of
CC hepatitis C virus, through the selection of those antibodies which bind
CC to the peptides. The peptides from hepatitis C virus hypervariable region
CC 1 of the envelope protein E2 are used to produce a medicament for raising
CC or increasing levels of antibodies able to bind HCV (hepatitis C virus)
CC HVR1 epitopes in a mammal. The medicament is used to treat or prevent an
CC HCV infection.
XX
SQ Sequence 27 AA;

Query Match 75.0%; Score 15; DB 21; Length 27;
Best Local Similarity 60.0%; Pred. No. 3e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 1 RTLSF 5

RESULT 4
AAE18094
ID AAE18094 standard; peptide; 27 AA.
XX
AC AAE18094;
XX
DT 07-MAY-2002 (first entry)
XX
DE Biotinylated hepatitis C virus region derived peptide (HVR), 272.
XX
KW Hepatitis C virus; HCV conjugate; immune response; therapeutic; virucide;
KW hepatotropic; antiinflammatory; HCV region derived peptide; HVR.
XX
OS Hepatitis C virus.
XX
PN WO200193804-A2.
XX
PD 13-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US17302.
XX
PR 02-JUN-2000; 2000US-209089P.
XX
PA (MERI) MERCK & CO INC.
XX
PI Conley AJ, Mckenna PM, Przysiecki CT, Keller PM;
XX
DR WPI; 2002-164292/21.
XX
PT Hepatitis C virus conjugate useful for inducing immune response in a
PT subject comprises a polypeptides or protein complex carrier and
PT immunogenic peptides covalently bonded to the carrier -
XX
PS Example 1; Page 22; 63pp; English.
XX
CC The patent discloses hepatitis C virus (HCV) conjugates able to induce
CC an immune response recognising different strains and variants of HCV.
CC The conjugates comprise a polypeptide or protein complex carrier and
CC one or more HCV mimotopes. Sequences of the invention are useful for
CC inducing an immune response in a subject e.g. human, chimpanzees, mice
CC or horses. They are also useful for the preparation of antisera, in
CC therapeutic/diagnostic applications to generate anti-HCV antibodies,
CC for detecting the presence of HCV in a subject and treating the subject
CC infected with HCV. The present sequence is biotinylated HCV region
CC derived peptide (HVR), 270.
XX
SQ Sequence 27 AA;

Query Match 75.0%; Score 15; DB 23; Length 27;
Best Local Similarity 60.0%; Pred. No. 3e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      4 RXLXF 8
          | | |
Db      1 RTLSF 5
```

RESULT 5
AAB38998
ID AAB38998 standard; Peptide; 28 AA.
XX
AC AAB38998;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human secreted peptide #20.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
PN WO200056880-A1.
XX
PD 28-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-US06781.
XX
PR 19-MAR-1999; 99US-0125363.
PR 08-DEC-1999; 99US-0169617.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-602220/57.
DR N-PSDB; AAC59706.
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating disorders such as Parkinson's and
PT Alzheimer's diseases, cancers and infections -
XX
PS Claim 11; Page 380; 422pp; English.
XX
CC Sequences AAB38971-B39020 represent the amino acid sequences of 50
CC human secreted proteins encoded by the genes AAC59679-C59728. The genes
CC and proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated from
CC a range of human tissues disclosed in the specification. The nucleic
CC acids, proteins, antibodies and (ant)agonists are useful in the
CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
SQ Sequence 28 AA;

Query Match 75.0%; Score 15; DB 21; Length 28;
Best Local Similarity 60.0%; Pred. No. 3.1e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      4 RXLXF 8
          | | |
Db      9 RTLAF 13
```

RESULT 6
AAY02875
ID AAY02875 standard; Protein; 30 AA.
XX
AC AAY02875;
XX
DT 11-JUN-1999 (first entry)
XX
DE Fragment of human secreted protein encoded by gene 85.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.

XX
PN WO9902546-A1.
XX
PD 21-JAN-1999.
XX
PF 07-JUL-1998; 98WO-US13684.
XX
PR 12-SEP-1997; 97US-0058785.
PR 08-JUL-1997; 97US-0051916.
PR 08-JUL-1997; 97US-0051918.
PR 08-JUL-1997; 97US-0051919.
PR 08-JUL-1997; 97US-0051920.
PR 08-JUL-1997; 97US-0051925.
PR 08-JUL-1997; 97US-0051926.
PR 08-JUL-1997; 97US-0051928.
PR 08-JUL-1997; 97US-0051929.
PR 08-JUL-1997; 97US-0051930.
PR 08-JUL-1997; 97US-0051931.
PR 08-JUL-1997; 97US-0051932.
PR 08-JUL-1997; 97US-0052732.
PR 08-JUL-1997; 97US-0052733.
PR 08-JUL-1997; 97US-0052793.
PR 08-JUL-1997; 97US-0052795.
PR 08-JUL-1997; 97US-0052803.
PR 18-AUG-1997; 97US-0055684.
PR 18-AUG-1997; 97US-0055722.
PR 18-AUG-1997; 97US-0055723.
PR 18-AUG-1997; 97US-0055947.
PR 18-AUG-1997; 97US-0055948.
PR 18-AUG-1997; 97US-0055949.
PR 18-AUG-1997; 97US-0055950.
PR 18-AUG-1997; 97US-0055953.
PR 18-AUG-1997; 97US-0055954.
PR 18-AUG-1997; 97US-0055964.
PR 18-AUG-1997; 97US-0055984.
PR 18-AUG-1997; 97US-0056360.
PR 12-SEP-1997; 97US-0058660.
PR 12-SEP-1997; 97US-0058661.
PR 12-SEP-1997; 97US-0058664.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Ebner R, Fischer CL, Kyaw H, Lafleur DW, Li Y, Moore PA;
PI Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR, Zeng Z;
XX
DR WPI; 1999-120770/10.
XX
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Disclosure; Page 101; 464pp; English.
XX
CC This sequence represents a fragment of a secreted human protein encoded
CC by the nucleic acid molecule detailed in the descriptor line. The gene
CC can be used to generate fusion proteins by linking to the gene to a
CC human immunoglobulin Fc portion (e.g. AAX27302) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 123 novel genes and their fragments (nucleic
CC acid sequences: AAX27311-X27449; amino acid sequences AAY02650-Y02788)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 123
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAX27311 for described uses).
XX
SQ Sequence 30 AA;

Query Match 75.0%; Score 15; DB 20; Length 30;
Best Local Similarity 60.0%; Pred. No. 3.4e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 12 RTLTF 16

RESULT 7
AAW45039
ID AAW45039 standard; peptide; 31 AA.
XX
AC AAW45039;
XX
DT 27-APR-1998 (first entry)
XX
DE Immunomodulatory peptide D22184AA.
XX
KW Immunomodulator; immunosuppressant; immunostimulator; treatment;
KW transplant rejection; autoimmune disease; cancer; infection.
XX
OS Synthetic.
XX
PN WO9739023-A1.
XX
PD 23-OCT-1997.
XX
PF 04-APR-1997; 97WO-SE00574.
XX
PR 23-SEP-1996; 96SE-0003469.
PR 12-APR-1996; 96SE-0001422.
XX
PA (ASTR) ASTRA AB.
XX
PI Bergstrand H, Eriksson T, Lindvall M, Saernstrand B;
XX
DR WPI; 1997-526397/48.
DR N-PSDB; AAV05459.
XX
PT Nucleic acids encoding cysteine- or methionine-containing peptide(s)
PT which have immuno:stimulatory or immunosuppressive activity - can be
PT used to treat, e.g. cancers, infection, auto:immune disease or
PT transplant rejection
XX
PS Claim 22; Page 162; 183pp; English.
XX
CC The present peptide is an immunosuppressant or immunostimulator. An
CC immunosuppressant can be used to treat transplant rejection or
CC autoimmune disease, e.g. rheumatoid arthritis, systemic lupus
CC erythematosis, Sjogren's syndrome, scleroderma, mixed connective
CC tissue disease, dermatomyositis, polymyositis, Reiter's syndrome,
CC Behcet's disease, type I diabetes, Hashimoto's thyroiditis, Graves'
CC disease, multiple sclerosis, myasthenia gravis, encephalomyelitis,
CC phemphigus vulgaris, vegetans or foliaceus, Senear-Usher syndrome
CC or Brazilian phemphigus. An immunostimulator can be used to treat
CC conditions such as cancer or infection.
XX
SQ Sequence 31 AA;

Query Match 75.0%; Score 15; DB 18; Length 31;
Best Local Similarity 60.0%; Pred. No. 3.5e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 23 RALAF 27

RESULT 8
AAY09486
ID AAY09486 standard; peptide; 31 AA.
XX
AC AAY09486;
XX
DT 14-JUL-1999 (first entry)

XX
DE Immunoactive peptide containing a mammalian signal peptide #1.
XX
KW Immunoactive; immunomodulation; immunosuppression; immunostimulation;
KW immune response; immunoreactive; autoimmune disease.
XX
OS Synthetic.
XX
PN WO9919347-A1.
XX
PD 22-APR-1999.
XX
PF 06-OCT-1998; 98WO-SE01801.
XX
PR 10-OCT-1997; 97US-0949024.
XX
PA (ASTR) ASTRA AB.
XX
PI Bergstrand H, Eriksson T, Lindvall M, Saernstrand B;
XX
DR WPI; 1999-287953/24.
XX
PT Synthetic genes encoding immunoreactive peptides containing cysteine
PT or methionine
XX
PS Disclosure; Page 12; 104pp; English.
XX
CC The present invention describes nucleic acid molecules comprising a
CC coding sequence encoding an immunoreactive peptide and further encoding
CC a protein targeting sequence. The nucleic acid is administered to a
CC patient so that its expression product, an immunoactive peptide,
CC modulates an immune response in a patient. The nucleic acid can also be
CC used to treat cancer, either after surgery to remove a portion of the
CC cancer or after ionizing radiation. A cytokine is also administered in
CC conjunction with the nucleic acid. Cells containing the nucleic acid
CC molecule can also be used for treatment. The immunoactive peptide is
CC immunosuppressive and can be used in patients with autoimmune disease.
CC The present sequence represents an immunoactive peptide from the
CC present invention.
XX
SQ Sequence 31 AA;

Query Match 75.0%; Score 15; DB 20; Length 31;
Best Local Similarity 60.0%; Pred. No. 3.5e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 23 RALAF 27

RESULT 9
AAB44770
ID AAB44770 standard; Protein; 33 AA.
XX
AC AAB44770;
XX
DT 12-FEB-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 9 SEQ ID NO:69.
XX
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; autoimmune disease; infection;
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW cerebrovascular disorder; nervous system disorder; ocular disorder;
KW wound healing; skin aging; food additive; preservative.
XX
OS Homo sapiens.
XX
PN WO200058336-A1.
XX
PD 05-OCT-2000.
XX
PF 23-MAR-2000; 2000WO-US07726.
XX
PR 26-MAR-1999; 99US-0126597.
PR 07-JAN-2000; 2000US-0174877.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-602355/57.
DR N-PSDB; AAC79807.
XX
PT Nucleic acid encoding human secreted proteins, used to treat, prevent,
PT ameliorate or diagnose medical conditions such as cancer, and
PT autoimmune diseases -
XX
PS Claim 11; Page 359; 391pp; English.
XX
CC The polynucleotide sequences given in AAC79799 to AAC79848 encode the
CC human secreted proteins given in AAB44762 to AAB44811. AAB44812 to
CC AAB44829 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are used in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities are:
CC immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
CC cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
CC neuroprotective; antibacterial; virucide; fungicide; and
CC ophthalmological. The polynucleotides and polypeptides can be used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases, hyperproliferative disorders, cardiovascular
CC disorders, cerebrovascular disorders, angiogenesis, nervous system
CC disorders, infections caused by bacteria, viruses and fungi and ocular
CC disorders. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. AAC79790 to AAC79798 and
CC AAB44761 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 33 AA;

Query Match 75.0%; Score 15; DB 21; Length 33;
Best Local Similarity 60.0%; Pred. No. 3.7e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 6 RSLSF 10

RESULT 10
AAM60971
ID AAM60971 standard; Protein; 33 AA.
XX
AC AAM60971;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 33076.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX

PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 33076; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 33 AA;

Query Match 75.0%; Score 15; DB 22; Length 33;
Best Local Similarity 60.0%; Pred. No. 3.7e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 14 RTLSF 18

RESULT 11
AAM32187
ID AAM32187 standard; Protein; 33 AA.
XX
AC AAM32187;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #6224 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 27; SEQ ID No 32456; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP:
CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
SQ Sequence 33 AA;

Query Match 75.0%; Score 15; DB 22; Length 33;
Best Local Similarity 60.0%; Pred. No. 3.7e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 14 RTLSF 18

RESULT 12
ABG41720
ID ABG41720 standard; Peptide; 33 AA.
XX
AC ABG41720;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human peptide encoded by genome-derived single exon probe SEQ ID 31385.
XX
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesis; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -

XX
PS Claim 27; SEQ ID No 31385; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomtosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a peptide/protein
CC encoded by a single exon probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 33 AA;

Query Match 75.0%; Score 15; DB 23; Length 33;
Best Local Similarity 60.0%; Pred. No. 3.7e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 RXLXF 8
| | |
Db 14 RTLSF 18

RESULT 13
ABG43538
ID ABG43538 standard; Peptide; 33 AA.
XX
AC ABG43538;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human peptide encoded by genome-derived single exon probe SEQ ID 33203.
XX
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesis; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
PS Claim 27; SEQ ID No 33203; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomtosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a peptide/protein
CC encoded by a single exon probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 33 AA;

Query Match 75.0%; Score 15; DB 23; Length 33;

Best Local Similarity 60.0%; Pred. No. 3.7e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      4 RXLXF 8
          | | |
Db     14 RTLSF 18
```

RESULT 14
AAB44347
ID AAB44347 standard; Protein; 35 AA.
XX
AC AAB44347;
XX
DT 14-FEB-2001 (first entry)
XX
DE Human secreted protein encoded by gene 13 clone HSREC72.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
PN WO200058358-A1.
XX
PD 05-OCT-2000.
XX
PF 23-MAR-2000; 2000WO-US07725.
XX
PR 26-MAR-1999; 99US-0126602.
PR 14-JAN-2000; 2000US-0176063.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-594640/56.
DR N-PSDB; AAC79009.
XX
PT Fourty nine nucleic acid molecules encoding human secreted proteins,
PT useful in the prevention, treatment and diagnosis of cancer, immune
PT disorders, cardiovascular disorders and neurological diseases -
XX
PS Claim 11; Page 342; 367pp; English.
XX
CC Sequences AAB44335-B44382 represent the amino acid sequences of 49
CC human secreted proteins encoded by the genes AAC69084-C69119. The genes
CC and proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated from
CC a range of human tissues disclosed in the specification. The nucleic
CC acids, proteins, antibodies and (ant)agonists are useful in the
CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
SQ Sequence 35 AA;

Query Match 75.0%; Score 15; DB 21; Length 35;
Best Local Similarity 60.0%; Pred. No. 3.9e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      4 RXLXF 8
          | | |
Db      9 RSLTF 13
```

RESULT 15
AAB63066
ID AAB63066 standard; Protein; 36 AA.
XX
AC AAB63066;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 18 SEQ ID NO:76.
XX
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; vulnerary; gene therapy; neoplasm;
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder; infection;
KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;
KW skin aging; food additive; preservative.
XX
OS Homo sapiens.
XX
PN WO200061748-A1.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US08982.
XX
PR 09-APR-1999; 99US-0128696.
PR 14-JAN-2000; 2000US-0176069.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-638566/61.
DR N-PSDB; AAF22333.
XX
PT New nucleic acid molecules encoding 48 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 11; Page 436; 480pp; English.
XX
CC AAF22316 to AAF22363 encode the human secreted proteins given in AAB63049
CC to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins
CC and polypeptides homologous to them. Human secreted proteins have
CC activities based on the tissues and cells the genes are expressed in.
CC Examples of activities include: immunosuppressive; antiarthritic;
CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
CC cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
CC fungicide; ophthalmological; and vulnerary. The polynucleotides and
CC proteins can be used to prevent, treat or ameliorate a medical condition
CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
CC sheep. They are also used in diagnosing a pathological condition or
CC susceptibility to a pathological condition. Disorders which are diagnosed
CC or treated include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. The polypeptides can also be
CC used to aid wound healing and epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities, fat content,
CC lipid, protein, carbohydrate, vitamins, minerals, cofactors and other
CC nutritional components. AAF22307 to AAF22315 and AAB63048 represent
CC sequences used in the exemplification of the present invention.

XX
SQ Sequence 36 AA;

Query Match 75.0%; Score 15; DB 21; Length 36;
Best Local Similarity 60.0%; Pred. No. 4e+02;
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       4 RXLXF 8
           | | |
Db      31 RSLAF 35
```

Search completed: December 14, 2002, 15:45:42
Job time : 59 secs